University of Alberta

Studies of Secondary Structural Propensities and Hydrophobic Clustering in the α-helical Model Coiled-Coils

by

Stanley Chun Ho Kwok

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

DOCTOR OF PHILOSOPHY

Department of Biochemistry

Edmonton, Alberta

SPRING, 2004

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-96289-X
Our file Notre référence
ISBN: 0-612-96289-X

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

The α-helical coiled-coil is a protein dimerization motif that participates in both structural and regulatory roles. This structure is defined by right-handed amphipathic α-helices which pack together via hydrophobic interactions, and provides an ideal model to study protein folding because its repetitive architecture maintains a well-defined protein core and also imparts other non-covalent interactions.

The sequence of a coiled-coil is characterized by the heptad repeat ***abcdefg*** where ***a*** and ***d*** are non-polar residues, that provide hydrophobic stabilization, but secondary structural propensities, ionic, clustering and other interactions also affect folding. Understanding the context-dependency of these stabilizing and destabilizing interactions remains a challenge, and deciphering how competing short- and long-range interactions affect coiled-coil conformation is our approach to understand protein folding.

We designed *de novo* coiled-coils where native secondary structural elements were inserted into different environments. The results illustrate that whereas some sequences are malleable and can be stabilized in multiple conformations, other sequences with structural determinants are very selective towards a defined conformation. These ideas improve our current understanding of secondary structure folding and conformational switches, as well as illustrate the dominance of short protein motifs.

Another context-dependence phenomenon is the arrangement of hydrophobes in long coiled-coils, where the protein core often incorporates small, non-polar residues despite their destabilizing characters. Our research confirmed that the distribution of non-polar stabilizing and other destabilizing residues is non-random in native coiled-coils, and these evolutionary adaptations allow flexibility in unstable regions for biomolecular recognition. In contrast, regions of high stability with hydrophobic clustering maintain structural integrity.

These investigations represent novel approaches to identify sequences critical for stability and function. Rather than studying protein folding one residue at a time, we described universal stabilization properties such as secondary structural specificity determinants and hydrophobic clustering that represent "words" or higher order information packets. Identifying these protein motifs is essential to distinguish sequences of functional significance from ambivalent sequences, thus minimizing extraneous sequence information. Not only do these examples illustrate the intricacies of context-dependent stabilization in secondary structures but, more importantly, it has universal relevance to protein fold prediction, conformation and stability and *de novo* design.

Dedicated to my Parents

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. Robert S. Hodges for being the driving force behind this thesis. Without his constant guidance and inspirations, this thesis would not have come to fruition

I would like to thank all members of the Hodges lab, past and present who have helped and worked with me over the last few years. In particular, Colin Mant, Stephen Lu, Paul Cachia, Brian Tripet, Kim Oikawa, Lorne Burke, Marc Genest, and Jennifer Labrecque. Thank you also to Morris Aarbo, Mundeep Chana, Jason Chen, Jason Moses, Rob Hodgins, Wayne Kohn, Masoud Jelokhani-Niaraki, Pierre Lavigne, Nancy Berton, Darin Lee, Anthony Mehok, Traian Popa, Paul Semchuk, Bob Luty and Kurt Wagschal - thank you for sharing your insights and enthusiasm for research.

I would like to thank the Alberta Peptide Institute for the infrastructure necessary to carry out efficient experiments, especially Jack Moore, Bob Parker and Mike Carpenter.

This work was generously supported by Canadian Institutes for Health Research and Protein Engineering Network Centres for Excellence. I would personally like to thank the Alberta Heritage Foundation for Medical Research and the Natural Sciences and Engineering Research Foundation for scholarships over the years.

TABLE OF CONTENTS

Page

Chapter I. Introduction..... 1

I. Introduction to protein folding..... 1

A. Emerging theme of protein folding: the energy landscape..... 5

B. Non-covalent interactions in protein folding..... 7

i) Hydrophobic interactions..... 8

ii) Packing and van der Waals interactions..... 10

iii) Polar and hydrogen bonding interactions..... 11

iv) Electrostatic interactions..... 13

v) Secondary structural propensities..... 15

vi) Capping, turns and other motifs..... 18

C. Model protein systems for studying non-covalent interactions. 19

i) α-Helical coiled-coils as models for the study of protein folding..... 21

D. Detail of the coiled-coil structure..... 24

i) The *a* and *d* positions of a coiled-coil form the hydrophobic core..... 24

ii) The *e* and *g* positions of a coiled-coil flank the hydrophobic core..... 26

iii) The *b*, *c* and *f* positions are solvent-exposed.....27

iv) Super-coiling, hydrophobic packing and oligomerization states.....27

v) Short coiled-coils versus long coiled-coils..... 30

E. Goals..... 31

Chapter II. Materials and Methods.....33

I. Materials.....33

A. Chemicals and reagents.....33

II. Methods.....35

A. Solid-phase peptide synthesis.....35

B. Purification of peptides - Reversed-phase high performance liquid chromatography..... 42

C. Peptide oxidation.....45

D. Compositional analyses: amino acid analysis and mass spectroscopy.....45

E. Size determination: size-exclusion chromatography and ultracentrifugation..... 47

F. Probe of secondary structure - circular dichroism spectroscopy..... 49

i) Temperature-induced denaturation monitored by circular dichroism..... 52

ii) Chemical denaturation monitored by circular dichroism.... 53

G. Determination of enthalpy change and heat capacity by calorimetry..... 55

Chapter III. Structural cassette mutagenesis in a *de novo* designed protein: proof of a novel concept for examining protein folding and stability..... 57

I. Introduction..... 57

II. Results and discussion..... 62

A. Structural cassette mutagenesis (SCM): approach and design... 63

B. SCM: initial objective.....65

C. Propensity models.....66

D. Design of cassette holder coiled-coil.....70

E. Selection of cassette for insertion into coiled-coil.....73

F. Helicity and stability of the native model coiled-coil.....80

G. Effect of increasing β-sheet propensity of the cassette on overall coiled-coil stability.....89

H. Effect of decreasing hydrophobicity of non-polar core of the cassette on overall coiled-coil stability.....95

III. Conclusions.....99

Chapter IV. Importance of secondary structural specificity determinants in protein folding: insertion of a native β-sheet sequence into an α-helical coiled-coil.....102

I. Introduction.....102

II. Results.....107

A. Design of the α-helical coiled-coil host protein as a cassette holder.....107

B. Selection of the parent cassette for insertion into the coiled-coil.....112

C. Characterization of the host coiled-coil following insertion of parent cassette.....114

D. Identification of potential determinants that prevent α-helix formation in the cassette.....118

E. Verification of the Asn-Asn motif as a helix destabilizing secondary structure specificity determinant.....121

F. Contribution of structural specificity determinant regions to coiled-coil stability.....122

III. Discussion..... 124

Chapter V. Clustering of large hydrophobes in the hydrophobic core of two-stranded coiled-coils controls protein folding and stability.... 133

I. Introduction..... 133

II. Results..... 137

A. Design of the α-helical coiled-coils with different hydrophobic clustering..... 137

B. Secondary structure characterization by circular dichroism... 138

C. Sedimentation equilibrium analyses of the oligomeric states of coiled-coils..... 142

D. Comparison of coiled-coil stabilities by thermal and urea denaturation..... 142

E. Differential scanning calorimetry of coiled-coils..... 144

III. Discussion..... 146

Chapter VI. Stabilizing and destabilizing clusters in the hydrophobic core of long two-stranded α-helical coiled-coils..... 153

I. Introduction..... 153

II. Results..... 157

A. Stabilizing and destabilizing clusters in natural coiled-coils. 158

B. Intervening regions (outside clusters) in natural coiled-coils. 165

C. Coiled-coil design..... 167

D. Structural and biophysical characterization of coiled-coils... 172

E. The effect of a three-residue alanine cluster on coiled-coil stability..... 179

F. The effect of Leu to Ala substitutions juxtaposed between a destabilizing cluster and a destabilizing residue..... 182

III. Discussion..... 184

Chapter VII. Effect of chain length on coiled-coil stability: decreasing stability with increasing chain length..... 192

I. Introduction..... 192

II. Results..... 196

A. Design of three α-helical coiled-coils with different stabilities..... 196

B. Structural and biophysical characterization of coiled-coils with different lengths..... 203

C. Successive insertion of helical heptads into the coiled-coil host..... 210

D. Helical heptad insertions into other model coiled-coils..... 216

III. Discussion..... 218

Chapter VIII. Future directions..... 224

I. Overview..... 224

A. Conversion to β-sheet amyloid structure..... 224

B. Temperature profiling to characterize hydrophobic clustering..... 227

II. Impact and outlook..... 229

Bibliography..... 232

Curriculum Vitae..... 260

LIST OF TABLES

Table		Page
2-1	List of reagents	33
3-1	α-Helical and β-sheet propensities of the twenty naturally occurring amino acids	69
3-2	Summary of results and HPLC retention data of oxidized and reduced coiled-coils	84
4-1	Summary of circular dichroism results	117
4-2	Change in free energy for different non-covalent interactions	126
4-3	Numerical analysis of secondary structural specificity determinants.	130
5-1	Biophysical characterization of reduced and oxidized hydrophobic clustered peptides	140
5-2	Thermodynamic parameters of clustered peptides	147
6-1	Analysis of clusters in myosin and tropomyosin	160
6-2a	Occurrence of clusters in the hydrophobic core of myosin	163
6-2b	Occurrence of clusters in the hydrophobic core of tropomyosin	164
6-3	Biophysical data for disulfide-bridged coiled-coil analogs	175
6-4	Context dependent changes in stability for Leu to Ala substitutions.	181
7-1	Biophysical data for disulfide-bridged coiled-coil analogs with β-heptad inserts	205
7-2	Stability data for disulfide-bridged coiled-coil analogs with β-heptad inserts	208

LIST OF FIGURES

Figure		Page
1-1	Schematic representation of protein folding	3
1-2	An energy landscape schematic for Acylphosphatase	6
1-3	Selected coiled-coil structures	23
1-4	Models of side chain packing at positions ***a*** and ***d*** of coiled-coil	29
2-1	Four general steps in solid-phase peptide synthesis	36
2-2	Deprotection of N-α-protection groups: *t*-Boc vs. Fmoc	38
2-3	Formation of activated amino acid-HOBt ester	39
2-4	HPLC separation of pure peptide from crude product	44
2-5	Sedimentation equilibrium analyses	50
3-1	Peptide and protein designs for propensity studies	68
3-2	Models of the structural cassette mutagenesis coiled-coil host	71
3-3	Ribbon drawings of λ light chain immunoglobulin	74
3-4	Secondary structural propensity score of immunoglobulin fragment	75
3-5	Helical wheel representation of cassette inserted into coiled-coil	77
3-6	Molecular surface model of β-sheet cassette in a-helical coiled-coil	79
3-7	Amino acid sequences of cassettes inserted into coiled-coil	81
3-8	Circular dichroism spectra of selected coiled-coils	83
3-9	GdnHCl Denaturation profile of selected coiled-coils	86
3-10	Stability contribution of successive Thr to Ala substitution	93

3-11 RP-HPLC and stability profiles of selected coiled-coils with different hydrophobic core substitutions..... 100

4-1 Model of the structural cassette mutagenesis coiled-coil..... 109

4-2 Molscript drawing of the λ light chain of immunoglobulin..... 113

4-3 Amino acid sequences of cassettes for insertion into coiled-coil..... 115

4-4 Circular dichroism spectra of selected coiled-coils..... 116

4-5 GdnHCl denaturation profiles of selected coiled-coils..... 120

5-1 Coiled-coil nomenclature, sequences and schematics of three- and two-cluster peptides..... 136

5-2 Biophysical characterization of coiled-coils..... 141

5-3 Stability comparison of three- and two-cluster peptides..... 143

5-4 Thermal melting and differential scanning calorimetry of three-cluster peptide..... 145

5-5 Schematic representation of the hydrophobic core residues of tropomyosin..... 150

6-1 Stabilizing and destabilizing clusters and intervening regions in human myosin and human tropomyosin..... 166

6-2 Peptide nomenclature, sequences and schematics with different destabilizing clusters..... 168

6-3 Models of coiled-coils with stabilizing and destabilizing cluster..... 173

6-4 Representative circular dichroism spectra of the most stable coiled-coil..... 174

6-5 Sedimentation equilibrium analyses of selected coiled-coils..... 178

6-6 Thermal denaturation profiles of selected coiled-coils..... 180

6-7 Percentile distribution of cluster size in stabilizing and destabilizing clusters in myosin and tropomyosin..... 185

7-1 Proposed relationship between stability and coiled-coil length.....195

7-2 Peptide nomenclature, sequences and schematics of coiled-coils with different length.....198

7-3 Helical coiled-coil representation of the cross-sections of the three heptads used in this study..... 199

7-4 Representative circular dichroism spectra of the shortest coiled-coil.. 204

7-5 Stability profile of selected coiled-coils: Urea and temperature melting..... 212

7-6 Correlation plot of coiled-coil chain length to thermal stability: oxidized versus reduced..... 213

7-7 Correlation plot of coiled-coil stabilities in three different models.....215

7-8 Comparison of protein stability from thermal melting and urea denaturation..... 217

8-1 Sequence comparison of parent peptide versus hydrophobic analog.. 226

8-2 Temperature profiling of coiled-coils..... 228

LIST OF ABBREVIATIONS

Amino acids:

Ala, A	alanine
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
Cys, C	cystein(e)
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
His, H	histidine
Ile, I	isoleucine
Leu, L	leucine
Lys, K	lysine
Met, M	methionine
Nle	norleucine
Phe, F	phenylalanine
Pro, P	proline
Ser, S	serine
Thr, T	threonine
Trp, W	tryptophan
Tyr, Y	tyrosine

Val, V	valine
Boc or *t*-Boc	*tert*-Butyloxycarbonyl
BOP	benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate
CD	circular dichroism
DCC	N,N′-dicyclohexylcarbodiimide
DCM	dichloromethane
DIC	1,3-diisopropylcarbodiimide
DIEA	N,N-diisopropylethylamine
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
DSC	differential scanning calorimetry
DTT	dithiothreitol
EDT	1,2-ethanedithiol
ESI	electrospray ionization
Fmoc	9-fluorenylmethoxycarbonyl
GdnHCl	Guanidinium hydrochloride
HBTU	N-[1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HF	hydrofluoric acid
HOBt	1-hydroxybenzotriazole
HPLC	high-performance liquid chromatography
MALDI-TOF	matrix-assisted laser desorption/ionization - time of flight
MBHA resin	methylbenzhydrylamine resin
NMM	N-methylmorpholine
NMP	1-methyl-2-pyrrolidinone

PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
Rink amide resin	4-(2′,4′-dimethoxyphenyl-Fmoc-aminomethyl)-phenoxy resin
RPC, RP-HPLC	reversed-phase high-performance liquid chromatography
SPPS	solid-phase peptide synthesis
TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
T_m	temperature denaturation midpoint
$[Urea]_{1/2}$	urea denaturation midpoint

CHAPTER 1

INTRODUCTION

This thesis describes research studies on protein folding and stability in *de novo* designed model coiled-coil proteins, and the applications of these results to understand better the effects of non-covalent and context-dependent interactions on protein conformation. The first part of the introduction is a summary of the up-to-date perspectives on protein folding with emphasis on hydrophobic interactions, secondary structural propensities and other non-covalent interactions. The introduction then moves onto a discussion of simple model proteins in studying structure and stability relationships, leading to the *de novo* protein design approach and the application of the two-stranded dimeric α-helical coiled-coil motif. Finally, the introduction addresses the specific topics outlined in the experiments that follow.

I) Introduction to protein folding

Proteins comprise the most important functional biological molecules because of their diverse roles as structural scaffolds, signaling hormones, cell-surface receptors, carriers and transporters, regulatory molecules and catalysts of biochemical reactions. The three-dimensional structure of a protein is crucial to the interactions with its environment, and the protein molecule must maintain specific

recognition surfaces and active sites with sufficient overall stability to interact with nucleic acids, lipids and other protein molecules to carry out chemical reactions, molecular communications and other processes necessary for life. The structure of a protein is generally encoded in its amino acid sequence, where the biological information encoded in the deoxyribonucleic acid (DNA) is translated to the one-dimensional arrangement of 20 naturally-occurring amino acids. The translation of the amino acid code to the final three-dimensional structure, i.e., the protein folding phenomenon, is still a great challenge to modern science, and unraveling this process will no doubt provide an unique insight into the way in which evolutionary selection has influenced the properties of biological molecules to optimize structure for function.

Excluding the proteins associated with biological membranes, there are two broad structural classes of proteins: the fibrous proteins, such as keratin and collagen, which are generally elongated molecules involved in structural scaffolds and the maintenance of cellular integrity and motility; and the globular proteins which are comprised of mostly cytosolic spherical molecules with compact hydrophobic cores such as lysozyme and myoglobin. The globular proteins are tightly folded polypeptide chains forming a hydrophobic core where non-polar residues are buried (Fig. 1-1). Historically, the protein folding problem tends to focus on the class of globular proteins because of their functional significance. Before any protein structures were visualized with modern spectroscopy techniques, protein chemists had long speculated that protein molecules would fold into highly

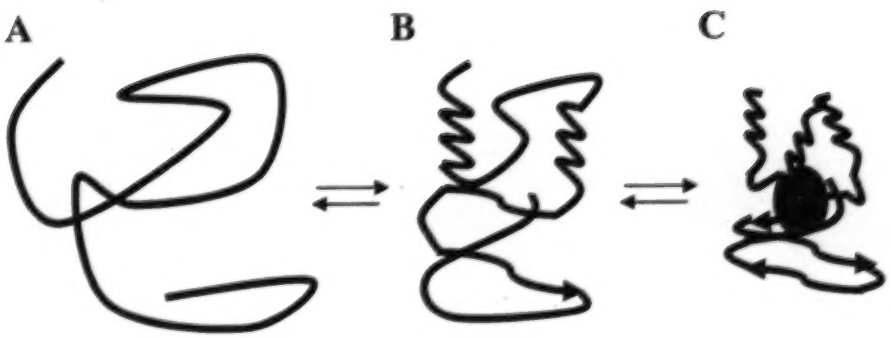


Fig. 1-1. Schematic representation of the folding of a globular protein in water. **A)** The extended polypeptide chain is directed towards hydrophobic collapse in an aqueous environment. **B)** The molten globule structure where non-polar amino acid side chains are oriented toward the inside of the protein with partial formation of secondary structures. **C)** The native three-dimensional conformation with a well-packed hydrophobic core (shaded grey), and the proper arrangements of secondary structures.

symmetrical, repeating polymers because of the repeating building units of the peptide backbone - consequently, it was postulated that protein structure and function could be easily understood. Not until the ground-breaking crystallographic structure determination of sperm whale myoglobin by John Kendrew and coworkers in 1958 did a much more complex picture of protein folding emerge (Kendrew *et al.*, 1958). Contrary to previous expectations, this first low-resolution protein structure showed that the myoglobin molecule adopted localized secondary structures such as α-helices and β-sheets, consequently forming larger complex asymmetric domains that illustrate the intricacies of protein architecture.

The fundamental proposed mechanism of protein folding is the thermodynamic hypothesis, i.e., the three-dimensional structure of the native protein almost always corresponds to the most thermodynamically stable conformation under physiological conditions. The pioneering work of Christian Anfinsen's group showed that protein folding is a self-assembly process; if a protein is denatured, it refolds back into shape with no other assistant molecules (Anfinsen, 1973). Thus, the logical hypothesis is that the conformation of a protein is determined by the specificities of all the non-covalent interactions within the protein and with its environment, which is mediated by the side chains of the constituent amino acids. Therefore, it is theoretically possible to predict the structure of a protein given its amino acid sequence; however, the total number of possible conformations of any polypeptide chain is so astronomically large, that an exhaustive search for this specific conformation would take an equally astronomical time. Consequently, the

protein folding problem illustrates this nearly impossible task of predicting the single native conformation with all the proper interactions within a polypeptide chain among the millions of possible structures from apparent random samplings. Yet in nature, small proteins can fold and refold rapidly *in vitro* in the aqueous environment, thus hinting that specific "pathways" from non-covalent interactions serve as guides to the final stable native folded state without random sampling of all possible conformations (Levinthal, 1969).

A) Emerging theme of protein folding: the energy landscape

The emerging theme to understand protein folding has evolved from the classical chemical pathways with mandatory steps between specific partially folded intermediates (Ptitsyn *et al*, 1973; Honig *et al.,* 1976; Karplus and Weaver, 1976) to a more stochastic simultaneous search of the many conformations accessible via multiple intermediates at lower energy states leading to the final most stable conformation (Dill *et al*, 1997; Shea and Brooks, 2001; Yon, 2002; Dobson, 2003). This statistical mechanic and polymer physics of the energy folding landscape is slowly supplanting the classical chemical dynamics description of protein folding because it explains the rapid folding of proteins *in vitro.* If the energy folding landscape describing the folding process has a generally sloping shape, only a small number of all possible conformations would be sampled by the polypeptide chain, thus speeding up the folding processes (Fig. 1-2). These sloping surfaces are guided

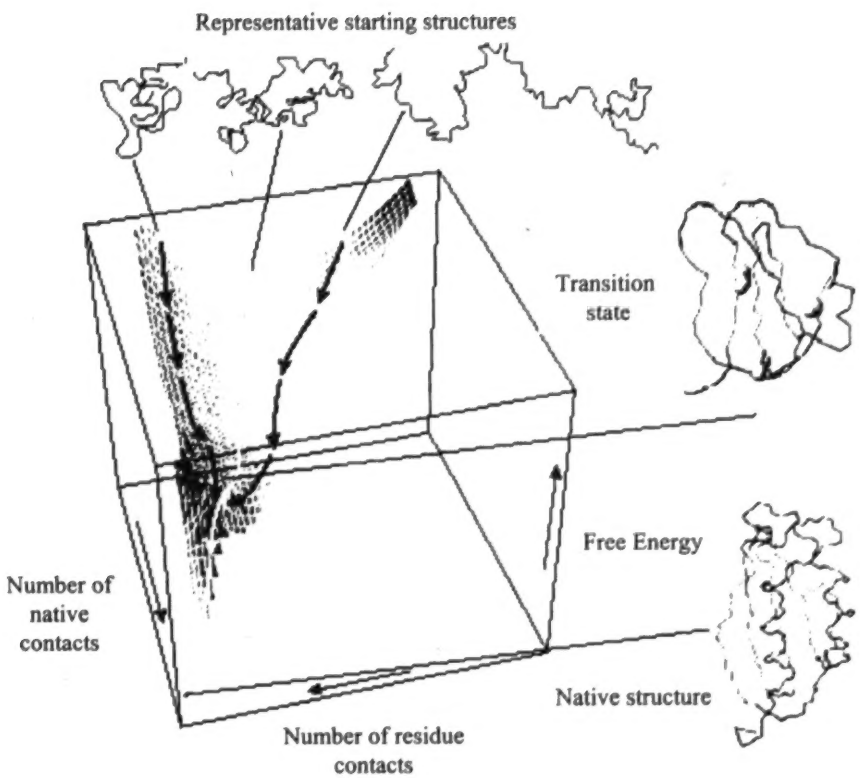


Fig. 1-2. An energy landscape schematic for the folding pathway of Acylphosphatase. Adapted from Vendruscolo *et al.* (2001).

by native-like interactions between residues because, on average, native-like interactions are more stable and thus more persistent than non-native ones, and progressive native-like contacts eventually lead to the lowest-energy conformation corresponding to native protein topology (Vendruscolo *et al.*, 2001).

One important question arises: how does the correct fold emerge from such fundamental steps? That is, how is the energy landscape unique to a specific protein defined by its amino acid sequence? Conceptually, the answer is addressed by the classical chemical dynamics (transition state) view of protein folding, because specific partially-folded intermediates and native-like residue contacts (which are measurable transition states) are still important limiting cases of the general protein folding mechanism. For example, the rapid formation of secondary structures such as autonomous α-helices and intrinsic β-turns occurs in the micro- and nano-second time frame, preceding the formation of the tertiary hydrophobic and electrostatic contacts (Ferguson & Fersht, 2003). Consequently, the investigation of specific non-covalent interactions in protein folding will continue to be crucial for investigating the most basic steps and determinants in protein folding

B) Non-covalent interactions in protein folding

Despite the sloping surface on an energy landscape representative of the free energy difference between the unfolded polypeptide chain and the biologically active native conformation, this net stability change in natural proteins is generally only 5-

10 kcal/mol (Creighton, 1990; Dill, 1990, Makhatadze and Privalov, 1995, Pace *et al.,* 1996). Although this difference is relatively small, it represents the sum of all the stabilizing and destabilizing non-covalent interactions that are much larger in magnitude. Stabilizing enthalpic interactions derived from residue-residue contacts contribute largely to protein stability, but these are offset by the loss of conformational entropy upon folding.

i) Hydrophobic interactions

Since the publication of Kauzmann's classic paper in 1959 detailing the significant role of hydrophobic interactions in specifying and stabilizing protein conformation, this driving force is thought to be the dominant non-covalent interaction promoting protein folding in an aqueous environment (Kauzmann, 1959). At room temperature in a benign chemical environment, the hydrophobic effect in proteins is largely-entropically driven, and occurs when water molecules arrange themselves around the hydrophobic side chains of non-polar amino acids in a rigid cage-like or clathrate structure that exhibits much higher organization than bulk water (high energy state) (Pertsemlidis *et al.,* 1996). This resultant decrease in solvent entropy is what drives the hydrophobic collapse of the non-polar amino acids, because the hydrophobic moieties are tightly-collapsed, removed from the aqueous environment (high energy state) and transferred into the relatively non-polar interior of the protein (low energy state). This effect is also important for other

biological processes such as the formation of lipid bilayers, membranes in organelles and cell boundaries, micelles, transport vesicles and other amphipathic lipid/protein complexes in the aqueous cellular environment.

The effect of ordering water molecules around the hydrophobic side chains of amino acids is a temperature-dependent thermodynamic process with both an enthalpic (ΔH) and an entropic (ΔS) component. As described earlier, protein folding is an entropically-favorable spontaneous process at room temperature, with a large positive partial heat capacity ΔC_p, because this value is higher in the multiple conformations of the unfolded states compared to the singular native structure. Although ΔC_p is dependent on the aqueous interactions of both hydrophobic and hydrophilic moieties of the polypeptide chain, the effect is largely attributed to the change in accessible hydrophobic surface area from non-polar amino acids, because high energy cost is associated with the transfer of non-polar compounds from a non-polar solvent to water (Nozaki and Tanford, 1971; Baldwin, 1986; Murphy *et al.*, 1990). The complexities of the "hydrophobic effect" phenomenon to describe protein folding originate from its temperature dependence. For example, high temperature-induced melting of protein structure is evidence that enthalpic factors can overcome unfavorable entropic factors of unraveling and exposing the hydrophobic core because the hydrophobic effect itself changes from being predominantly entropy-driven at low temperature to enthalpy-driven at high temperature (Makhatadze & Privalov, 1995).

A large literature of hydrophobicity scales has been published that estimates the net hydrophobicity of amino acid side chains and their contributions to overall protein stability (Fauchere and Pliska, 1983; Eisenberg and McLachlan, 1986; Sereda *et al.*, 1994). The experimental methods include the measurement of partition coefficients between an organic solvent and aqueous environment, the theoretical calculation of free energy values based on relative differences in solvent accessible surface area, the measurement of RP-HPLC retention behavior in small molecules, and the effects of site-directed mutagenesis on the stability of model proteins and peptides; overall, reviewers have pointed out that most estimates are in relative agreement with one another (Biswas *et al.*, 2003). The general classification is that Ile, Met, Leu, Val are the most hydrophobic aliphatic residues, and residues such as the aromatic Phe, Tyr, and Trp exhibit both polar and non-polar character depending on their environment. Estimates from modeling experiments, computer calculations and calorimetric and stability analyses of protein analogs have shown that a single buried methyl group ($-CH_3$) can contribute 0.1 – 1.5 kcal/mol to the overall stability relative to the hydrogen (-H) side-chain (Serrano *et al*, 1993; Jelesarov and Bosshard, 1996).

ii) Packing and van der Waals interactions

Site-directed mutagenesis experiments in globular proteins revealed that the proximal packing interactions from neighboring side chains may play a significant

role in protein stabilization, an equivalence to lattice energies in crystalline molecules. For example, in T4 lysozyme, mutational studies in which residues with large non-polar side chains such as Ile and Leu were substituted with small residues such as Ala and Val, resulted in destabilization that was much greater than could be explained by differences in hydrophobicity (Matthews, 1995; Xu *et al.*, 1998). In addition, the overall entropy of the whole protein, measured by the crystallographic thermal factor, increases in the analogs where large non-polar interior residues are replaced by smaller residues. Therefore, these enthalpically-stabilizing interactions arise from optimal packing between the aliphatic side chains. These van der Waals interactions, arising from London dispersion forces from temporary uneven distributions in electron charge density (induced dipoles), though small in magnitude, can add up significantly over the entire length of the protein molecule. Taken together, both hydrophobic interactions and packing interactions can have a significant effect on overall protein and conformation.

iii) Polar and hydrogen bonding interactions

Polar interactions in proteins are dominated by the sharing of hydrogen atoms between two electronegative atoms, specifically the hydrogen bonding interactions between peptide amide-carbonyl backbone and the side chain moieties with hydrogen bonding potentials. The protein backbone plays a major role in organizing secondary structures such as α-helices, β-sheets and turns because the amide

hydrogen of one amide group and the carbonyl of another can participate in the stabilization of secondary structures via hydrogen bonding. In the early days of protein folding studies, these interactions were thought to be the driving force of protein folding (because of the prominence of repeating secondary structures), but its perceived relevance was diminished as the hydrophobic collapse model and hydrophobic interactions were demonstrated experimentally to be more important (Pace, 2001).

Today, a common consensus among structural biologists is that the interactions of polar groups generally contribute little to the overall stability of proteins because the energetic cost of desolvation upon folding (removal of hydrogen bonding from the aqueous environment) is more than offset by the stabilizing energy of interactions between these groups (protein internal hydrogen bonding), i.e., inter-residue hydrogen bonds have to compete with the potential of hydrogen bonding with the bulk water (Honig and Yang, 1995; Lazaridis *et al.,* 1995). In other words, in the unfolded state, hydrogen bonding partners can interact with the aqueous environment; thus, there is no net increase in stability compared to hydrogen bonding in the folded conformation. In fact, it was estimated that each main chain hydrogen bond only contributes between 0.6-0.9 kcal/mol to the stability of an α-helix (Scholtz *et al.,* 1991a; Serrano, 2000). Despite these observations, approximately 70% of all amide backbone hydrogen bonding interactions in a globular protein are buried in the interior and out of contact with the aqueous environment (Pace *et al.,* 1996). Some structural biologists suggest that hydrogen bonds in the folded conformation could in

fact be much stronger than hydrogen bonds between residues in the unfolded state and bulk water because **1)** intra-residue hydrogen bonding may have a higher stability because of decreased entropic cost from residues of close proximity; **2)** hydrogen bonding within a secondary structural element comprises a cooperative network of interactions that add synergistically to overall stability; and **3)** hydrogen bonding stabilizing is highly context-dependent, where placing an -OH group at a site designed for a $-CH_3$ group is very unfavorable; and vice versa, the placement of -OH in a site that was selected for $-CH_3$ would be destabilizing despite the potential for hydrogen bonding. Consequently, a number of researchers now believe hydrogen bonding may play an equally important role in protein stabilization, and at the very least, this interaction alleviates the energetic cost of burying two polar groups in the partially hydrophobic protein environment (Creighton, 1990; Makhatadze & Privalov, 1995; Pace *et al.*, 1996).

iv) Electrostatic interactions

The two broad classifications of electrostatic interactions are non-specific long- range interactions between domains with a predominant ionic character; and specific, shorter-range interactions between the side chains of oppositely-charged residues. Mundeep Chana and coworkers have demonstrated the role of long-range electrostatics as a driving force to steer and orient large biological molecules prior to their specific binding via hydrophobic interactions (Chana *et al.*, 2002). More

significant for protein folding are the specific electrostatic interactions mediated by the charged groups on the side chains of amino acids: Arg, Asp, Glu, His and Lys, the free amine and carboxyl termini or groups, and other covalently-attached charged moieties such as phosphates or metal ions cofactors such as calcium and zinc. Although Coulomb's laws can be applicable to the description of these interactions, shortcomings lie in the approximation of point charges (dimensionless), whereas the electrostatics in a protein molecule occur at atomic dimensions. Furthermore, the charge potential on an amino side chain is dependent on the local protein environment (dielectric constant) and is sometimes delocalized within a group of atoms, e.g. the imidazole ring of His and the guanidinium group of Arg (Schutz & Warshel, 2001). Thus, relative to van der Waals interactions, electrostatic interactions are difficult to model because of their strong dependence on distance, solvent-protein environment and the geometry and character of the charged groups involved. However, their significance to protein stabilization is obvious by the relative increase of stabilizing electrostatic interactions in proteins derived from thermophilic organisms compared to their mesophilic counterparts (Zhou and Dong, 2003).

A common example of ion pairing in proteins occurs at a preferred distance of 3Å between the hydrogen bond acceptor ($-NH_3^+$ of Lys) and the donor atom ($-CO_2^-$ of Glu), and is referred to as a salt bridge (Barlow and Thornton, 1983). Unlike hydrophobic interactions and hydrogen bonding which are generally stabilizing, electrostatic repulsions between like-charged groups can destabilize protein

structure. Quantitatively, salt bridges on the surface of proteins appear to contribute only 0.2-0.5 kcal of net stabilization, presumably due to screening by bulk water (Dao-pin *et al.*, 1991; Horovitz *et al.*, 1990; Lyu *et al.*, 1992; Zhou *et al.*, 1994) but a single repulsion between like charges has been estimated to contribute 0.4-0.8 kcal/mol of net destabilization (Serrano *et al.*, 1990; Kohn, 1998). Interestingly, short-range electrostatic attractions outnumber electrostatic repulsions at a 3:1 ratio in globular proteins; also, 37% of these stabilizing ion pairings occur as part of a network (Barlow and Thornton, 1983). Perhaps in a similar manner to hydrogen bonding, the cooperativity of electrostatic interactions can synergistically enhance stabilizing potential. Therefore, the balance of electrostatic interactions from attractive and repulsive forces provide the necessary specificity to guide a final protein fold that should maximize stabilizing attractive forces but minimize unfavorable repulsions.

v) Secondary structural propensities

Secondary structural propensity is the measurement of the intrinsic potential of an amino acid to stabilize a defined secondary structure. Although these thermodynamic values are highly context-dependent, they provide a crude prediction of protein structure and stability based on amino acid sequence alone. The first propensity scale was actually a statistical frequency survey showing how certain amino acid residues occurred in α-helices (Chou and Fasman, 1978). Subsequent

surveys into the distribution of amino acids in β-sheets and β-turns showed that only a moderate amount of success can be achieved in secondary structural prediction by the summation of localized secondary structural propensities within a sequence. More recently, measurement of α-helical and β-sheet propensities has been extended to the determination of the effect of substituting different residues on the folding of a secondary structure element within a host peptide or in a guest site in a host protein (Hodges, 1996; O'Neil and Degrado, 1990; Chakrabartty *et al.*, 1991; Gans *et al.*, 1991). Although these propensity values are only modest in magnitude, the sum of local propensities within a sequence can have a dramatic effect on local stability, thus confirming the relevance of the early statistical survey of occurrence of amino acids in particular structural elements. However, because of the short-range nature of these interactions, secondary structure prediction based on propensity algorithms is generally correct for only 75% of the sequence (Eisenhaber *et al.*, 1996).

The recent emergence of protein misfolding diseases, i.e., protein conformation switching phenomena (Prusiner, 1998), has once again brought the interplay between α-helices and β-sheets into the limelight, thus leading to further investigations of intrinsic secondary structural propensities. In the Janus protein fold transformation project, researchers have successfully designed a stable, native-like helical protein that is 50% identical in sequence to a predominantly beta-sheet protein, the B1 domain of Streptococcal IgG-binding protein G (Dalal *et al.*, 1997). The properties of this protein illustrate the extent to which protein stability and conformation can be modulated through careful manipulation of key amino acid

residues, and have implications for understanding conformational change phenomena and in probing the malleability of the sequence/structure relationship. In a statistical survey of the Brookhaven Protein Data Bank, the predicted probability of finding a 6-residue or longer "chameleon sequence" approaches zero (0.021), but however, 38 sequences were found (Mezei, 1998). The longest such sequence consist of eight residues (Sudarsanum, 1998), and the existence of these sequences of non-trivial length confirms the structural adaptability of short sequences.

The physical basis of the observed difference in the secondary structure propensity of the naturally-occurring amino acid residues appears to be dependent on side chain entropy and its effect on the solvation of the polypeptide backbone. For example, the α-helical backbone restricts the number of rotamers that can be adopted by the side chains, thus reducing overall entropy. Furthermore, non-polar side chains in α-helices can also block access of bulk water to the hydrogen-bonded helical backbone. Thus, residues with non-polar side chains such as Ala, Met and Leu have high helical propensities. Because of its small size, Ala is the only residue that does not have reduced accessible rotamers in an α-helical conformation; thus, it has the highest intrinsic helical propensity value (Creamer & Rose, 1994). In contrast, polar and charged side chains are destabilizing because they compete with the protein backbone for hydrogen bonding partners; thus, they have lower helical propensities and higher sheet propensities. Furthermore, the β-branched side chains such as Thr and Val are specifically restructured by steric hindrance to the helical backbone and

showed high sheet preference. Despite these distinct classifications of residues as having high helical propensities or high sheet propensities, long-range interactions, in particular hydrophobic interactions, can often overcome the structural preference derived from intrinsic propensities because the burial of non polar surface area is still the greatest contributor to the stability of the folded conformation. Two chapters of this thesis (III and IV) will focus on the importance of intrinsic secondary structural propensities for the modulation of overall secondary structure in the context of an α-helical coiled-coil fold. Because of the limitation of intrinsic propensity being a localized effect, i.e. the single residual helical and sheet propensity do not affect long-range protein structure significantly, the natural extension of propensity study is the characterization of di, tri or tetrapeptide sequences with specific conformation preferences such as capping and turn motifs.

vi) Capping, turns and other motifs

Investigations on the orientations of side chain carbonyl and amino groups in high-resolution protein structure reveal that a common hydrogen bonding pattern is the main-chain backbone bonding to Asp, Asn, Thr and Ser, four to five residues ahead in the sequence (Wan and White, 1999a,b). These hydrogen bonding motifs occur in many secondary structure contexts, including Asx N-cap motif, β-turn, ST-motif and β-bulge, and are thought to have an inherent propensity as autonomous folding subunits. Among these different short sequence motifs, helix-capping motifs

are the most common and occur at or near the ends of helices, mediated by specific hydrogen bonding and hydrophobic interactions, because the first four N-H groups and last four carbonyl groups in a helix necessarily lack intra-helical hydrogen bonds. For example, Tripet and Hodges (2002) had recently revealed a set of strong intra- and inter-chain interactions in the neck region of the kinesin coiled-coil involving a helix capping box termed the "hydrophobic staple,". The thermodynamic contributions of these motifs may be small because of the short-range nature of the localized interactions; however, such "secondary structural determinants" may affect the organization of secondary structures. For example, Asx-motifs also occur in functionally interesting junctions in aspartyl proteases, EF hands, haemogloblins and glutathione reductases. Thus, one of the forthcoming research chapters addresses the importance of the Asn-Asn (NN motifs) in the disruption of helical coiled-coil conformation.

C) Model protein systems for studying non-covalent interactions

A large number of native globular proteins have served as models for studies of protein folding and protein-protein interactions since the introduction of site-directed mutagenesis to design specific mutants to probe the determinants of residual structure and function relationships (Matthews, 1995; Ohmura *et al.,* 2001; Minor & Kim, 1994; Munoz *et al.,* 1996; Pace *et al.,* 1996 and references therein). While these studies have achieved significant progress, they are limited by the large size of

the molecules (and possible complication of highly populated intermediates) and the complex context-dependence of intra-molecular interactions in a globular multiple-domains protein setting. From these limitations, a minimalist approach to understand protein folding was developed, i.e., a simpler model that contains fewer non-covalent interactions but focused enough to provide sufficient information in a well-defined protein environment. These models, particularly the application of peptides that contain <100 residues, make it easier to isolate and study specific non-covalent interactions while still maintaining a protein environment, thus increasing the likelihood that the results are comparable to native biological molecules (Hodges, 1996; Monera *et al*, 1995; Xiong *et al.*, 1995).

Protein and peptide *de novo* design is an arena that applies the experimental knowledge of the fundamental principles of protein folding, and has achieved remarkable success in the past few decades, building from the extensive literature since the early days of protein folding research. In a sense, the *de novo* design approach addresses the reverse protein folding problem - given the model of a particular protein conformation, predict what sequences would lead to the desired structure, rather than structural prediction from amino acid sequence. The lofty goal of *de novo* design is to incorporate different functions and structural elements from existing proteins and systematically build new biological molecules with new functions, activities and structures. To this end, a quantitative understanding of the relative and absolute contributions of all the different non-covalent interactions must be achieved.

i) α-Helical coiled-coils as models for the study of protein folding

In our laboratory, α-helical coiled-coil and bundle proteins have been some of the most successful targets of *de novo* design efforts. These molecules have several advantages which make them ideal for protein design. First, because of their repeating heptads and linearity, the design of α-helices reduces the protein folding problem to a one-dimensional problem from the three-dimensional globular protein context. There is only one type of secondary structure with a simple repeating architecture which is internally hydrogen bonded, forming an autonomous folding unit. In contrast, β-sheet proteins often require inter-domain hydrogen bonding and stabilization of other secondary structure such as turns and loops (Minor & Kim, 1994; Smith *et al.,* 1994). Consequently, the lack of other secondary structures allowed the experimenters to study the co-operative nature of the coiled-coil folding by simple biophysical techniques such as circular dichroism, micro-calorimetry and sedimentation equilibrium. Furthermore, α-helical coiled-coils are some of the simplest molecules containing long-range tertiary interactions, forming a stable well-packed hydrophobic core which models the interior of native proteins (Zhu *et al.,* 1993; Behe *et al.,* 1991; Alber, 1992; Harbury *et al.,* 1993). Because of the variation in orientations, packing, and oligomerization states, the careful design of coiled-coil proteins can accurately mimic both the short- and long-range non-covalent interactions found in all natural proteins (Kohn & Hodges, 1998 and references therein).

Before the visualization of any high-resolution structure, Crick predicted the two-stranded, repeating nature of α-helical coiled-coils in 1953 based on X-ray diffraction patterns (Crick, 1953). The two-stranded parallel α-helical coiled-coil, which contains just two interacting helices, is the simplest dimerization motif in nature. Such motifs are extremely abundant and are estimated to occur in 3% of all know protein sequences and play diverse biological roles such as transcription factors, chaperones, muscle molecules and viral fusion proteins (Fig. 1-3). Thus studying structure, function and stability interactions in the coiled-coil has applications for increased understanding of biological systems. Added to these complexities are higher-order helical bundle proteins that form trimers and tetramers. For example, laminin and fibrinogen are trimeric coiled-coils involved in cytoskeletons and transport, while Rop is a tetrameric coiled-coil membrane receptor (Engel, 1991; Timpl & Brown, 1994; Regan & Jackson, 2003). In addition, five-, six- and seven- transmembrane helical proteins have also been observed (Cohen & Parry, 1990; Kamtekar & Hecht, 1995).

Functionally, a subclass of short coiled-coils serves as important regulators of gene transcription. The so-called "leucine zippers" refer to a class of transcription factors with <50-residue coiled-coil domains such as GCN4, c-Myc, c-Max, Jun and Fos that can homo- and hetero-dimerize with different specificities to activate RNA synthesis (Alber, 1992). In conjunction with other coiled-coil regions found in the transcription protein motifs such as basic-helix-loop-helix or zinc-finger motifs, these molecules show the diverse roles of coiled-coils in nature. Besides their

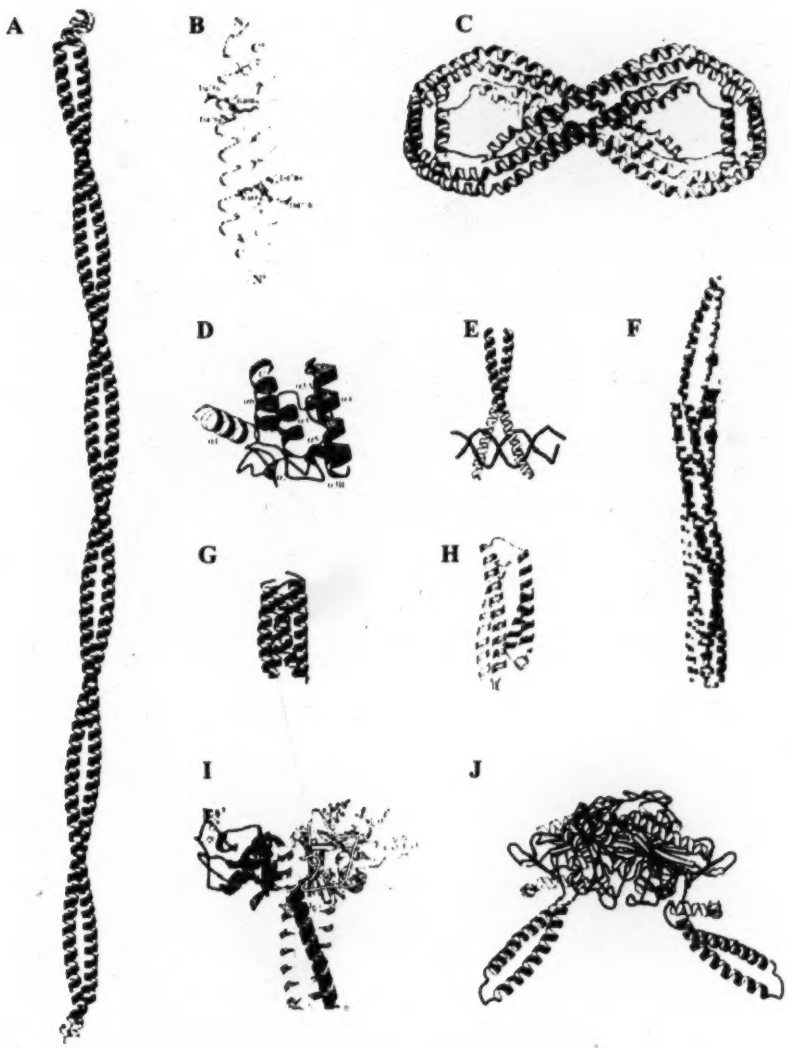


Fig. 1-3. Selected coiled-coil structures illustrating diversity. **A)** Tropomyosin **B)** Osmosensor Prop **C)** Apolipoprotein A1 **D)** Calponin homology domain **E)** GCN4 **F)** Serine chemotaxis receptor **G)** ROP **H)** Apolipoprotein E3 **I)** Lung surfactant protein **J)** Seryl tRNA-synthase. Figures were adapted from Lupas *et al.* (1996), Gruber and Lupas (2003), Zoetewey *et al.* (2003), and Shrive *et al.* (2003).

regulatory roles, coiled-coil motifs are intricately linked to the cytoskeleton and motility proteins (Singh & Hitchcock-DeGregori, 2003; Sheetz, 1999; Barr, 2000). The two-stranded tropomyosin molecule was the first coiled-coil sequenced, and it mediates the calcium-dependent movement of thin and thick filaments in muscle contraction (Hodges, 1972; Perry, 2001) Other coiled-coils serve as higher order structural scaffolds such as laminin, dynamin, kinesin and microtubule molecules.

More than two decades of coiled-coil research has yielded an extraordinary amount of information on this structure, and a detailed model of many coiled-coil non-covalent interactions has now emerged that offers excellent potential for *de novo* design. The investigations of non-covalent interactions in coiled-coils have significant relevance to the general phenomena of protein stability, structure and function, and represent the bulk of this thesis.

D) Detail of the coiled-coil structure

All the non-covalent interactions that control protein folding (hydrophobic, hydrogen bonding, electrostatic, secondary structural propensities and specific motifs) are readily apparent in the structure of the coiled-coil proteins.

i) The *a* and *d* positions of a coiled-coil form the hydrophobic core

The major determinant of coiled-coil conformation is adequate amphipathicity arising from correct arrangement of hydrophobic and hydrophilic

resides. The canonical α-helix has a periodicity of about 3.6 residues per turn or 7.2 residues per two turns, and therefore the spacing of hydrophobic residues 3- or 4-residues apart would result in an amphipathic helix. This arrangement would allow the burial of non-polar side chains upon the dimerization of amphipathic helices, forming an optimal hydrophobic core. The coiled-coil is therefore characterized by a seven-residue (heptad) repeat designated ***abcdefg,*** where positions ***a*** and ***d*** are usually non-polar residues. This 3-4 or 4-3 repeat of hydrophobes (**HXXHXXXHXXH**……. where H is a non-polar residue) was first revealed by the sequencing of the first coiled-coil molecule tropomyosin (described above), in which 71 of the 82 ***a*** and ***d*** positions contain residues with some non-polar character (Hodges *et al.,* 1972). These hydrophobic interactions are the major determinants of coiled-coil stability, and numerous site-directed mutagenesis studies have shown that the replacement of a large hydrophobic core residue in the core of a coiled-coil with Ala is extremely destabilizing (Zhou *et al.,* 1992a; Zhu *et al.,* 1993; Wagshcal *et al.,* 1999; Tripet *et al.,* 2000). Recent papers by Tripet and Wagschal, where the authors substituted all 20 naturally occurring amino acids except Cys (Orn) into both ***a*** or ***d*** positions of short coiled-coils, produced proteins with varying stabilities, up to a difference of 7.4 kcal/mol (Wagschal *et al.,* 1999; Tripet *et al.*,2000). In general, their results showed good but not exact correlation of protein stability to side chain hydrophobicity and, therefore, other factors such as packing and model-dependent context interactions must also be considered (Zhou *et al.,* 1993; Gonzalez *et al.,* 1996; Liu *et al.,* 2002). In nature, short coiled-coils tend to have a

very non-polar hydrophobic core, intermediate length coiled-coils can accommodate some charged and polar residues and long (>200-residue) coiled-coils often have up to 40% non-polar residues in the core (Sellers, 2000; Brown *et al.*, 2001). Therefore, it is of interest to investigate how non-polar residues are distributed and clustered in the different biological settings.

ii) The *e* and *g* positions of a coiled-coil flank the hydrophobic core

The ***e*** and ***g*** positions of the coiled-coil flank the hydrophobic core and are at the hydrophobic core/cellular environment interface and have potential interactions with both non-polar residues at ***a*** and ***d*** positions and any aqueous environment. In a parallel coiled-coil, a residue at position ***g*** of one strand is opposite the ***e*** heptad position in the other helix, thus allowing for interactions across the dimer interface (***i to i'+5)***. Concomitant with their inter-helical interactions, the ***e*** and ***g*** residues can also shield the hydrophobic core from entropically-unfavorable contacts with the bulk-solvent and make direct stabilizing van der Waals contacts that increase overall protein stability. Furthermore, the placement of bulky hydrophobes in these positions enlarges the hydrophobic face of the amphipathic helices and leads to higher-order helical bundle proteins. In native two-stranded parallel coiled-coils, these positions are populated by charged residues such as Lys, Arg, Glu, Asp, and His that have salt-bridge forming potential, often forming a network of inter-chain electrostatics that mediates fold nucleation, maintains helix registry and increases

overall protein stability. The doctoral thesis of Wayne Kohn is an excellent review of significant research in the elucidation of inter-helical ionic interactions at these positions (Kohn, 1998).

iii) The *b*, *c* and *f* positions are solvent-exposed

These heptad positions are solvent-exposed and primarily uninvolved in inter-helical contacts of two-stranded coiled-coils; however, they can still exert an influence on stability via intrinsic helical propensities and allow the coiled-coils to interact with other biological molecules. In the larger four-stranded helical bundles, the ***b*** and ***c*** residues can contribute some interactions across the helix interface (Betz *et al.,* 1997; Hill *et al.,* 2000). Tropomyosin is an example of a two-stranded coiled-coil protein where these solvent-exposed positions mediate the interactions with actin and the troponin complex to facilitate the sliding movement responsible for muscle contraction, and therefore the identities of these residues are important for recognition and binding (Brown *et al.,* 2001; Hitchcock-DeGregori *et al.,* 2002).

iv) Super-coiling, hydrophobic packing and oligomerization states

Because the α-helix has a periodicity of 3.6 residues per turn, for the hydrophobic faces of the two α-helices to interact in a coiled-coil, a gradual twist around the helix circumference is necessary. Crick proposed that the amphipathic

helices of a coiled-coil would wind around one another like strands of a rope in a left-handed fashion in order to maintain contacts between the 3-4 repeating non-polar residues (Crick, 1953). This supercoiling of the helical strand results in the narrower periodicity of 3.5 residues per turn observed in coiled-coils, thus permitting the hydrophobic side chains of one helix to pack "knobs-into-holes" into its neighbor. As a consequence of this asymmetric curvature, the backbone hydrogen bonding is shorter in the hydrophobic face and longer in the hydrophilic face. The coiled-coil helices cross at about 18° and the pitch of the supercoiling varies depending on the oligomerization states, ranging from a narrow 150 Å for a dimer to a larger 200 Å for a tetramer (Kamtekar & Hecht, 1995; Seo & Cohen, 1993).

As a general consensus, the oligomerization behavior of a given coiled-coil is largely affected by the packing arrangements of the hydrophobic residues at ***a*** and ***d*** positions in the core (Wagschal *et al.*, 1999; Tripet *et al.*, 2000). Analyses of high-resolution coiled-coil crystal structures revealed the asymmetric packing in the two non-identical core positions (Fig. 1-4). In position ***a***, the side chain C_α-C_β bond vector points away from the helix-helix interface, parallel to the C_α-C_α bond vector of the "hole" of the opposing helix, and is designated as parallel packing. In contrast, the C_α-C_β bond vector of ***d*** residue points into the interface, making a perpendicular angle with the C_α-C_α bond of the neighboring helical strand; thus, it is designated as perpendicular packing. Consequently, β-branched non-polar residues such as Ile and Val are favored at position ***a*** and unfavorable in positions ***d*** because of steric hindrance (Zhu *et al.*, 1993). Although some polar or charged residues are tolerated

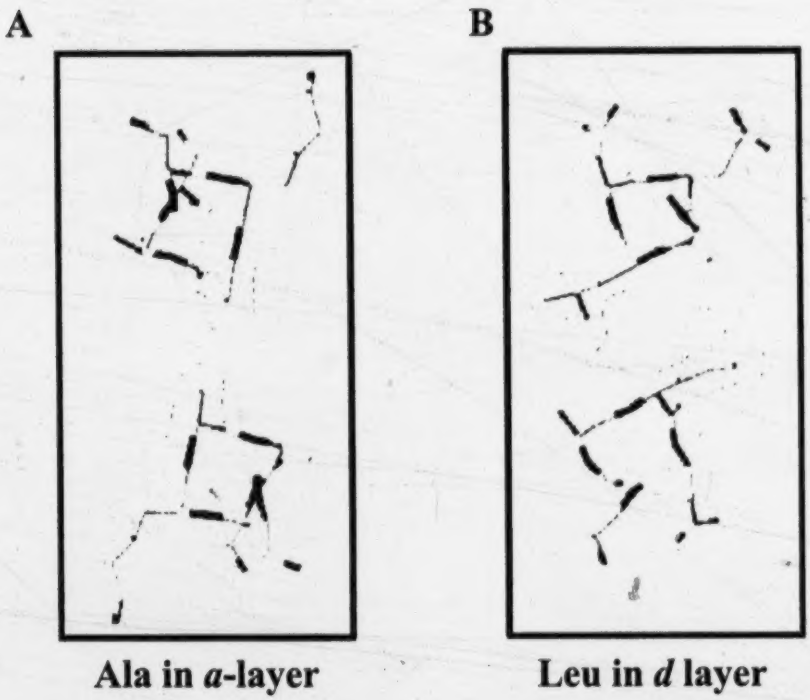


Fig. 1-4. Models of alanine and leucine side chain packing at positions ***a*** and ***d*** of a parallel coiled-coil, respectively. **A)** The side chains of Ala show parallel packing in the ***a***-layer. **B)** The side chains of leucine show perpendicular packing in the ***d***-layer.

in both core positions, their roles seem to involve regulating chain orientation and homo- vs. hetero- dimerization rather than influence of overall oligomerization states. Further description of distribution of both non-polar and polar residues will be described in the experimental section.

v) Short coiled-coils versus long coiled-coils

Native coiled-coil domains can be classified into two general categories: **1)** the short (<50 residue) dimerization motifs, such as GCN4; and **2)** the longer coiled-coil molecules such as myosin and tropomyosin. As mentioned earlier, short coiled-coils tend to have extremely uniform and non-polar hydrophobic protein cores. In short coiled-coils, folding occurs relatively rapidly in a two-state approximation with no detectable intermediates (Holtzer *et al.,* 1997); furthermore, widespread helical secondary structure is generally not present at the rate-limiting step of hydrophobic collapse (Jelesarov & Bosshard, 1996). In unfolding experiments, these short proteins tend to fray and unfold cooperatively from the ends of the helices from the lack of backbone hydrogen partners, as observed by NMR experiments (Holtzer *et al.,* 1997). Furthermore, the unfolded state of short coiled-coils is predominantly the plethora of random arrangements know as random coil.

In contrast, the folding of long coiled-coils (> 200 residues), such as the tropomyosin molecule, has been shown to involve multi-state kinetics and intermediates composed of both rapid association and multiple slower rearrangement

steps (Paulucci *et al.*, 2002; Singh & Hitchcock-DeGregori, 2003). It is not uncommon for longer coiled-coils to exhibit biphasic denaturation and melting profiles indicative of distinct regions of high and low stability; not an unusual observation considering that independent subdomains within such a long coiled-coil structure can form autonomous folding units. Recently, Silva and co-workers observed that tropomyosin dissociates in a "knobs in a string" arrangement where discrete cooperative blocks unfold before the two strands separate (Suarez *et al.*, 2001). These observations culminate in a general puzzling phenomenon: why do fragments of long coiled-coils not fold despite the presence of repeating amphipathic heptads? Therefore, investigation into the stability densities within a coiled-coil would be beneficial for the understanding of different design paradigms in short and long coiled-coils.

E) Goals

The goal of this thesis, in general terms, is the understanding of non-covalent amino acid interactions important for protein folding in the coiled-coil model. We hope to discover and characterize interesting interactions using the *de novo* designed peptide approach with the following foci:

1) A quantitative understanding of some of the non-covalent interactions previously observed in specific non-ideal coiled-coil models. The

importance lies in not only how these interactions affect protein stability, but also how short sequences can modulate protein folds despite their localized contributions. These studies would expand previous research from our and other laboratories, and provide predictive generalizations relevant to coiled-coil protein folding.

2) The feasibility to design coiled-coils with desired folding propensity and characteristics under solution conditions that model interactions in biological sequences, with emphasis on short secondary structural elements, short hydrogen bonding motifs, chain length effects and hydrophobic and destabilization clustering phenomena in long coiled-coils.

3) The collection of these data in predictive algorithms to find and elucidate potential coiled-coil sequences, that would accurately identify both short and long coiled-coil domains and characterize regions of significant and marginal stabilities with consideration of hydrophobic and propensities effects.

CHAPTER II

MATERIALS AND METHODS

I) MATERIALS

A) Chemical and reagents

Unless otherwise stated, all chemicals and solvents used were analytical grade. Most of the chemicals used in this research are available from common chemical sources (Fisher, Aldrich, VWR International, Pierce), so suppliers are only listed for specialty chemicals and selective reagents (Table II-1). All solutions were prepared with deionizied water (>17.5MΩ) treated through a Milligen Milli-Q purification system. DIEA is distilled over ninhydrin, prior to use in peptide synthesis.

Table 2-1 List of Reagents

Reagent	Supplier
acetonitrile, HPLC grade	Fisher Scientific (Fairlawn, NJ)
anisole	Aldrich Chemical Co. (Milwaukee, WI)
t-Boc protected amino acids	Novabiochem (La Jolla, CA)

Fmoc protected amino acids	Novabiochem (La Jolla, CA)
BOP	Novabiochem (La Jolla, CA)
DCM, bulk	Fisher Scientific (Fairlawn, NJ)
DIC (solution)	Advanced Chemtech (Louisville, KY)
DIEA	Sigma Chemical Co. (St. Louis, MO)
DMF	Fisher Scientific (Fairlawn, NJ)
DMSO, HPLC grade	Aldrich Chemical Co. (Milwaukee, WI)
DTT	Aldrich Chemical Co. (Milwaukee, WI)
EDT	Aldrich Chemical Co. (Milwaukee, WI)
GdnHCl, electrophoresis grade	Fisher Scientific (Fairlawn, NJ)
HBTU	Novabiochem (La Jolla, CA)
HF	Matheson Tri-Gas (Newark, CA)
HOBt	Novabiochem (La Jolla, CA)
MBHA resin	Novabiochem (La Jolla, CA)
Ninhydrin, sigma grade	Sigma Chemical Co. (St. Louis, MO)
NMM, peptide synthesis grade	Fisher Scientific (Fairlawn, NJ)
Piperidine, 99.0%	Sigma Chemical Co. (St. Louis, MO)
PyBOP	Novabiochem (La Jolla, CA)
Rink Amide MBHA resin	Novabiochem (La Jolla, CA)
TFA	Halocarbon Products Corp. (River Edge, NJ)
TFE, 99+ %	Sigma Chemical Co. (St. Louis, MO)
Trypsin, sequencing grade	Promega (Madison, WI)

II) METHODS

A) Solid phase peptide synthesis

Solid-phase peptide synthesis was developed by R. Bruce Merrifield in the early 1960s, and represented an innovation in organic chemistry for the synthesis of small organic molecules (for review see Bodanszky, 1993, 1998; Fields, 1997; Merrifield, 1997). The advantage of solid-phase synthesis is the retention of the desired product on an insoluble support matrix that allows the drainage of excess unreacted reagents and undesirable side products. For this monumental achievement, Merrifield was awarded the Nobel Prize in Chemistry in 1984, and the solid-phase technique has been widely used in all applications in organic synthesis, including carbohydrate synthesis, peptidomimetic synthesis and combinatorial libraries.

Fig. 2-1 illustrates the four general steps in a SPPS scheme: **1)** anchoring, **2)** deprotection **3)** coupling and **4)** cleavage. SPPS is always carried out in the C- to N-terminal direction and anchoring is the coupling of the first N^{α}-protected (C-terminal) amino acid onto the solid support, usually via nucleophilic attack of a free amine on the linker on the solid support on the activated carboxyl group of the first amino acid (Fig. 2-1A). The efficacy at this step is crucial to ensure high yield of peptide product and is usually involved in an overnight reaction to ensure reaction to completion.

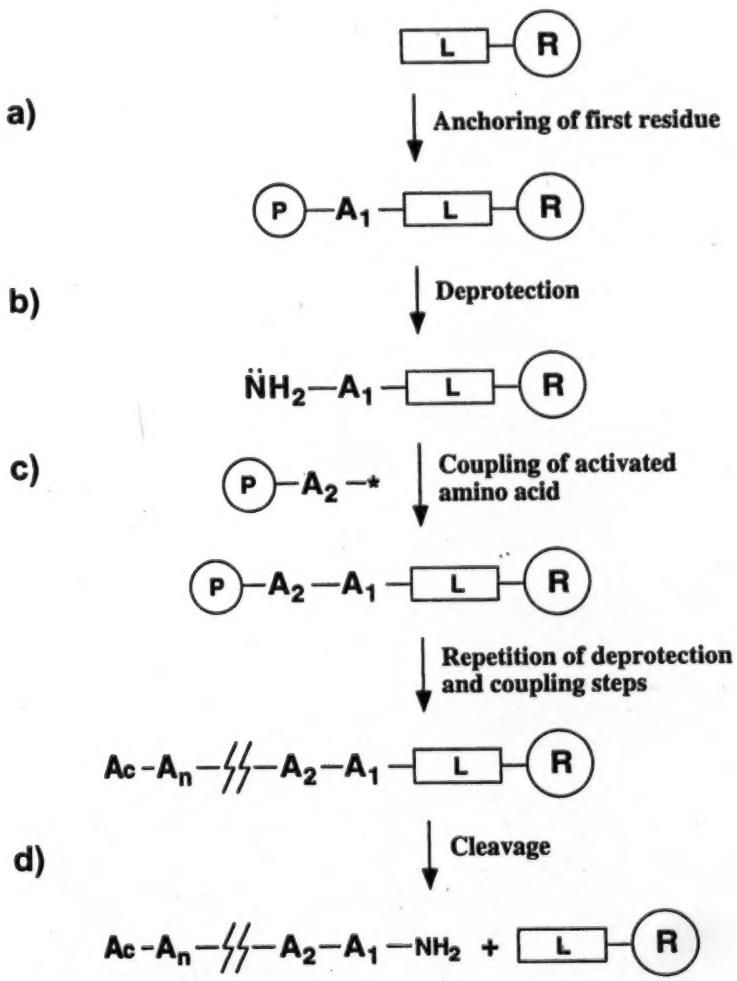


Fig. 2-1. Four general steps in a SPPS scheme: 1) anchoring of first amino acid residue; 2) deprotection of N-α-protecing group; 3) coupling of activated amino acid and 4) cleavage from the solid support.

The next step in peptide synthesis is the removal of the N^{α}-protection group on the attached amino acid to generate a free N^{α}-amine for the next nucleophilic attack leading to the formation of a peptide bond (Fig. 2-1B). There are two conventional N^{α}-protecting strategies commonly used: *t*-Boc, representing acid-labile chemistry, and Fmoc, representing base-labile chemistry (Fig. 2-2). *t*-Boc is removed with 50% TFA in DCM, followed by neturalization with a tertiary amine, for example, DIEA, and Fmoc is removed with a secondary amine, for example, 20% piperidine in DCM.

Following deprotection, the next N^{α}-protected amino acid is activated to provide an excellent leaving group for the nucleophilic attack by the free amine immobilized on the resin. A wide variety of chemical approaches is available for the *in situ* and pre-activation of the carboxyl moiety of the incoming amino acids (Albericio and Carpino, 1997; Novabiochem, 2003). The formation of activated HOBt esters is an effective strategy, and it is usually mediated by phosphonium-activating reagents (for examples, HBTU, BOP, PyBOP) which, in the presence of a hindered tertiary base (DIEA or NMM), convert the protected amino acid into an HOBt ester intermediate (Fig 2-3). The peptide bond is then formed by nucleophilic attack upon this ester intermediate, thus increasing the nascent peptide chain by one residue. Subsequent repetition of deprotection and coupling steps, with drainage of excess reagents and byproducts after each cycle, allows for stepwise elongation of the peptide chain. The key to a successful synthesis is extreme efficiency at each chemical step. For example, to synthesize a 39-residue polypeptide, a moderate

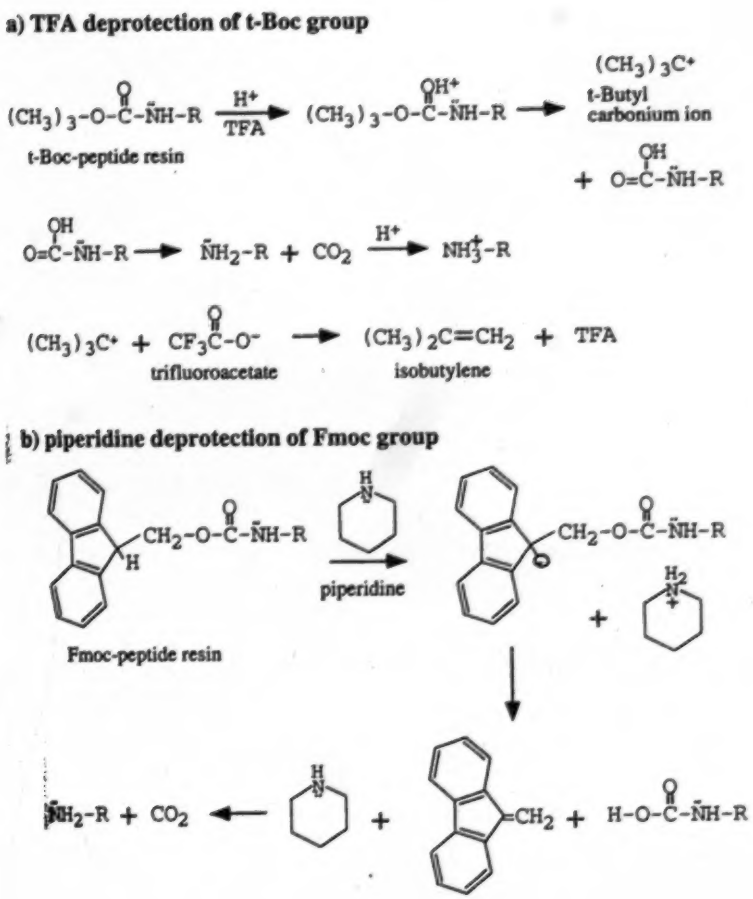


Fig. 2-2. *t*-Boc and Fmoc N-α-protecting strategies and deprotection schemes. **a)** *t*-Boc deprotection strategy via acid-catalyzed reactions by TFA; **b)** Fmoc deprotection via base-catalyzed chemistry via piperidine.

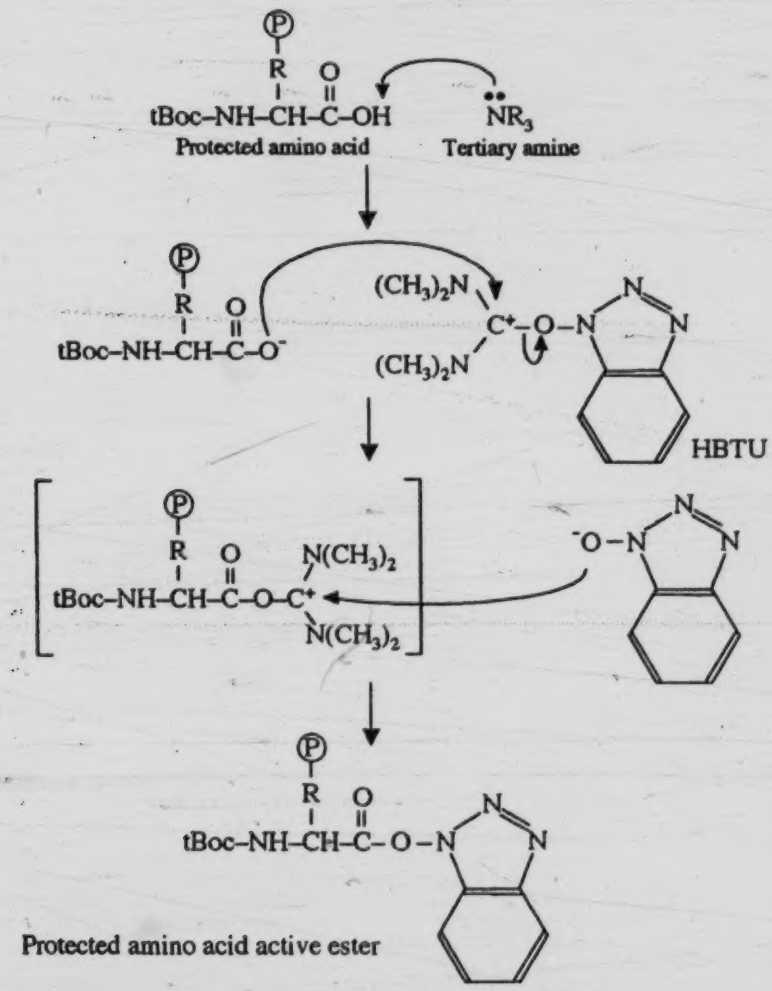


Fig. 2-3. Formation of N-α-protected but activated amino acid ester via base-catalyzed HBTU chemistry. The tertiary amine extracts a proton from the carboxylic acid moiety of the incoming amino acids. Subsequent nucleophilic attacks on HBTU and rearrangements generated the activated amino acid ester.

coupling efficiency of 99.0% (with the assumption of complete deprotection), would result in a final peptide yield of 67.6% ($99.0\%^{n}$ where n=39). Thus, one goal of successful peptide synthesis is to drive the reaction to extreme efficiency (>99.9%). Careful monitoring after each coupling by Kaiser test (Novabiochem, 2003) and repeat coupling at difficult sequences is employed to maximize reaction completion.

The final step of SPPS is cleavage from the peptidyl-resin. Once the peptide chain is completed, the final N^{α}-protected amino acid is deprotected, and then it is usually acetylated by reaction with acetic anhydride. The nascent peptide chain is then released from the resin using chemistry that is dependent on the lability of the resin linker. For a *t*-Boc synthesized peptide on a MBHA linker, strong acid anhydrous HF sealed in a resistant apparatus that may not readily available to many researchers is required for cleavage. In contrast, cleavage of Fmoc synthesized peptide on a Rink amide linker can be done with the milder acid TFA under conventional laboratory conditions. In both cases, the use of scavenging agents is encouraged to prevent the "reattachment" of deprotected groups on functionalized amino acids from adding back on the peptide backbone irreversibly and non-specifically. The use of anisole is a common strategy and EDT is also used to prevent thiol oxidation in peptides containing Cys (as do the peptides in the current studies).

All the peptides in this study were synthesized by the solid-phase methodologies either manually in a solid-phase extraction vessel, or on an Applied Biosystems automated peptide synthesizer Model 430A or 431A using small scale

0.1 mmol cycles. The general approach in solid-phase methodology uses conventional *t*-Boc chemistry, where the nascent polypeptide chain is synthesized on copoly(styrene, 1% divinylbenzene)-4-methylbenzhydrylamine-HCl (MBHA) resin, 100-200 mesh, substitution 0.73 mmol amino groups per gram with appropriately side-chain protected amino acids. Polystyrene MBHA resin is a favored polymer for research batch synthesis because of a large loading capacity (0.3 - 1.6 mmol/g) with excellent compatibility with a wide range of organic solvents, including DCM, DMF, NMP and DMA. Activation reagent O-benzotriazol-1-yl-1,1,3,3 tetramethyluronium hexafluorophosphate (HBTU, 0.45 M) and anti-racemization reagent N-hydroxylbenzotriazole (HOBt, 0.45M) were dissolved in DCM and DMF (20:80). DMF is an excellent solvent for the reactive species and intermediates, and DCM swells the polystyrene support to minimize steric hindrance. Coupling time was 60 min to 90 min with mild shaking agitation, and progress of coupling was monitored by Kaiser-ninhydrin test.

A major obstacle in coupling amino acids onto the nascent peptide is incomplete reaction due to aggregation from hydrophobic interactions or backbone hydrogen bonding, leading to the undesirable formation of protein secondary structure. This phenomenon usually occurs between 10-15 residues from the C-terminus and can lead to a significant decrease in peptide yield. To overcome this difficulty, double coupling with additional solvent additives such as 15% DMSO, NMP, NMM and/or the alternative use of activation reagents such as PyBOP and BOP are employed.

During the deprotection step, the removal of the *t*-Boc group from the resin was achieved with mild agitation in 50% TFA in DCM for 10 minutes. Excess TFA was drained and washed away by multiple washes in DCM, and the peptidyl-resin was neutralized by immersion in 10% DIEA in DCM. Once all cycles of coupling and deprotection are completed, the final completed peptide product was released from the MBHA resin support via hydrogen fluoride cleavage (HF) with anisole and EDTA. HF cleavage is generally performed at -4°C to prevent side reactions, and the reaction times range from 60 to 90 minutes depending on the character of the protecting groups on the amino acid side chains. The released product was then extracted from the resin with glacial acetic acid and dried overnight via lyophilization to yield the crude peptide product.

Recently, our laboratory has achieved better success with DIC as activation reagent, but the drawback of DIC is a longer reaction time in the deprotection and coupling steps relative to the conventional *t*-Boc and HBTU chemistry. However, because synthesis with DIC does not require base-activation (DIEA or NMM), the final product is of high purity with no racemized side products.

B) Purification of peptide - Reversed-phase high performance liquid chromatography

Reversed-phase HPLC (RP-HPLC) is a fast separation technique for peptide purification with high efficiency, resolution and recovery rate (Mant & Hodges, 2002 and references therein). Separations in RP-HPLC are based on overall

hydrophobicity, as the interactions between the non-polar side chains, the mobile phase (organic modifier) and the stationary phase (alkyl chains) denature the biomolecules. Mobile phases are generally mixtures of aqueous organic solvent, for example, methanol, acetonitrile, isopropanol, and the molecules are eluted off the columns by an increasing gradient of this organic modifier. A typical RP-HPLC column is made of porous silica beads derivatized with alkyl chains ranging from C8 to C18. Additional bulky and hydrophobic diisopropryl and diisobutyl groups are also covalently linked to the organo-silane to stabilize the solid support matrix and prevent column breakdown. However, the operation of RP-HPLC is still kept at low pH via ion-pairing reagents (e.g., TFA or formic acid) to suppress undesirable ionization of silanol groups which may lead to non-specific binding to basic moieties. As well, ion-pairing reagents also serve to increase retention time and minimize peak broadening because of improved interactions between the peptides and the stationary phase. Furthermore, the maintenance of low pH also minimizes the dissolvation of silanol groups, which are unstable in basic conditions.

Fig. 2-4 shows the typical purification profile of **1)** a crude product and **2)** the purified peptide. Typically, the crude peptides were purified by reversed-phase chromatography (RP-HPLC) on a Zorbax semi-preparative SB-C8 column (250 x 9.4 mm I.D., 5 μm particle size, 300 Å pore size) by a linear AB gradient rate of 0.2% B per minute, where eluent A is 0.05% aqueous TFA and eluent B is 0.05% TFA in acetonitrile with a flow rate of 3 ml/min (reviewed by Mant *et al.*, 1997b,c). For extremely difficult purification schemes, high temperature (70 °C) was used to

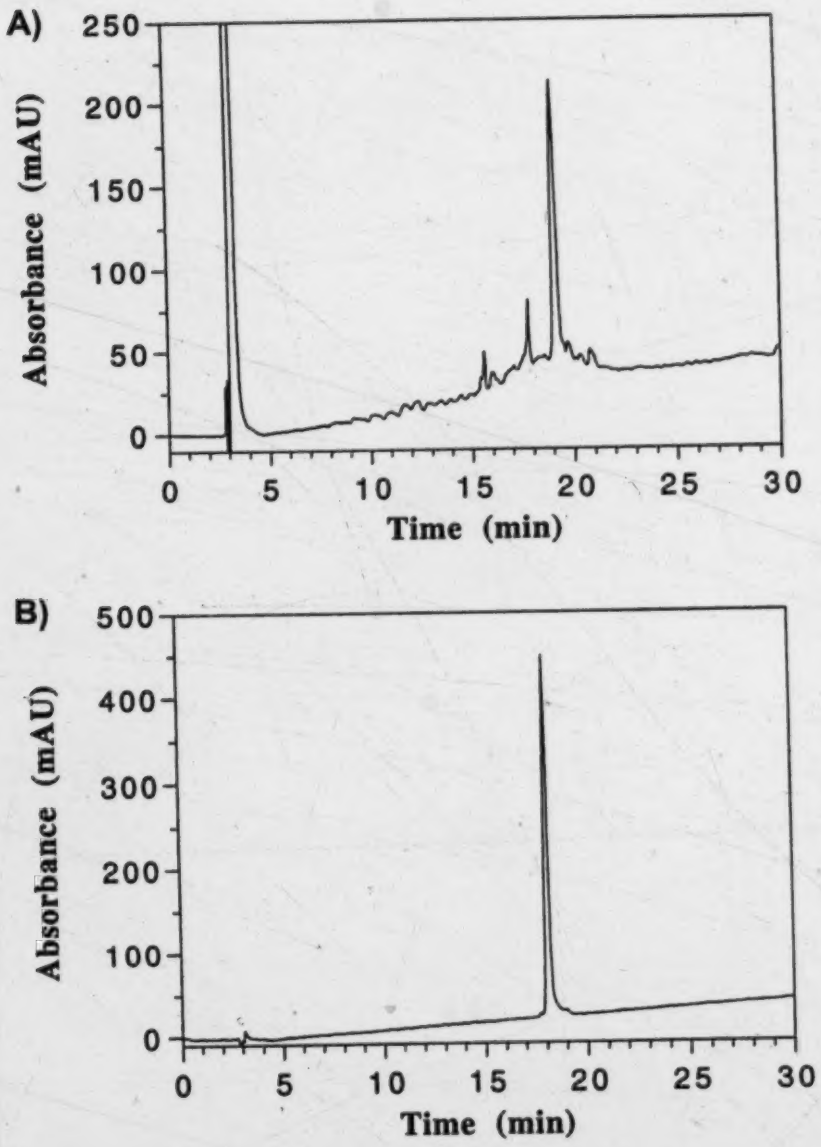


Fig. 2-4. HPLC separation of pure peptide from crude product. **A)** RP-HPLC analysis of crude peptide C-G-G-K-E-V-E-A-L-K-A-Q-V-T-S-A-T-L-A-I-T-G-L-Q-A-E-V-E-A-L-K-K-E-V-E-A-L-K–amide in a 2% increasing acetonitrile elution. **B)** Pure peptide in after a 0.2% increasing acetonitrile gradient in preparative RP-HPLC.

enhance column resolution of impurities. Analytical RP-HPLC was carried out to ensure purity and homogeneity of the peptide on a 4.6 mm I.D. x 150 mm Zorbax SB-C8 column using a linear AB gradient rate of 1% B per minute with the above eluents, and the final product was verified by quantitative amino acid analysis (Beckman Model 6300 amino acid analyzer), and by electrospray mass spectroscopy on a Fisons Quattro (Fisons, Pointe-Claire, Quebec, Canada).

C) Peptide oxidation

Formation of the disulfide-bridged homo-two-stranded coiled-coil was obtained by overnight stirring in a 100 mM NH_4HCO_3 buffer, pH 8.5, and the progress of the reaction was monitored by RP-HPLC. Typical peptide concentration during oxidation ranges from 5-10 mg/mL to favour the formation of peptide dimer. Upon satisfactory yield of the oxidized product, the reaction was neutralized with 5% aqueous solution of acetic acid, and the reaction product was purified by RP-HPLC. In the rare instance where the oxidized and reduced peptides are co-eluted, the unreacted reduced peptides were first derivatized with N-ethyl maleimide to add hydrophobicity and increase retention time relative to the oxidized product. The final mass of the oxidized coiled-coil was confirmed by mass spectroscopy.

D) Compositional analyses: amino acid analysis and mass spectroscopy

As an internal standard, a quantitative amount of Nle (5 nmol) was added to the peptide samples to determine sample loss during acid hydrolysis. Purified

peptides and proteins were hydrolyzed in 6.0 N HCl with 0.1% phenol at 160°C for 2 hr or at 120°C for 20 hr in evacuated, sealed glass tubes. After hydrolysis, HCl was removed by vacuum or rotary evaporation in the presence of KOH pellets, and the dried constituent amino acids were re-suspended in aqueous sodium diluent buffer (2% sodium citrate, 0.7% hydrogen chloride and 0.1% phenol, pH 2.2). The sample is then spun to remove particulates and the supernatant is injected onto a Beckman model 6300 amino acid analyzer with post column ninhydrin derivatization detection. The analyzer runs via a cation-exchange mechanism, and a typical analysis is 90 minutes long, consisting of three isocratic gradient steps with increasing pH from the starting condition of pH 3.0 up to the final pH 6.0 buffer with concurrent increases in operating temperature (45-65°C).

Molecular masses of crude or purified peptides and proteins were confirmed by electrospray ionization (ESI) and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy. ESI is particularly favorable for analyses of RP-HPLC samples because of aqueous sampling, but MALDI-TOF provides an alternative for short peptide samples with few charge moieties. Mass spectroscopy was carried out in positive mode to detect positively-charged ions and the typical purified RP-HPLC sample was diluted to 1 μM and dissolved in 50% aqueous acetonitrile with 0.1% formic acid or TFA prior to injection to enhance ionization. Once a sample is detected by the instrument, multiple spectra were acquired and accumulated (15-20 runs) to produce the raw data. First, the baseline of the raw data spectra is subtracted, then the data peaks are centered and

mathematically smoothed, and finally the determination of the transformed molecular weight is carried out by selecting adjacent ion peaks for the series that deconvolute to expected mass. Both ESI and MALDI-TOF suffered from a marked decrease in sensitivity when contaminating salts, small organic molecules and detergents are present; in particular, the carried-over anisole and EDTA scavengers during peptide cleavage especially hinder mass detection via the ESI technique. Therefore, the MALDI-TOF approach is a good alternative for identifying crude peptide mass because this analysis requires co-crystallization of samples with an acidified matrix before ionization and is less susceptible to contaminant interference.

E) Size determination: size-exclusion chromatography and ultracentrifugation

The effective molecular weight of peptide and proteins can be estimated by size-exclusion chromatography (SEC) from a standard curve. SEC was carried out on a Superdex 75 column (290 mm x 13 mm I.D., 10 μm particle size, 125 Å pore) with a flow rate of 0.3 ml/min. Typical operating condition is elution with isocratic benign solvent of aqueous buffer with 50mM PO_4, 500mM KCl, pH 7.0 at room temperature. The high salt concentration was used to prevent non-specific ionic interactions between the sample and the column matrix. The elution time is compared to a standard curve constructed with peptides of known molecular weights and no aggregative behavior, to provide a crude estimate of apparent molecular weight and oligomerization tendency (Mant *et al.*, 1997a).

The SEC results were confirmed by sedimentation equilibrium analysis which was performed on a Beckman XLA analytical ultracentrifuge with absorbance optics at 215 nm for the detection of peptide backbone or at 274 nm for the detection of tyrosine. Samples were first dialyzed exhaustively against an aqueous solution of 100mM KCl, 50mM PO_4, pH 7.0 (benign buffer) at 4°C. A 100µl aliquot was then loaded into the 12 mm, Epon cell (charcoal-filled), and the initial loading concentrations of peptide stock solutions ranged from 50 – 500 µM in benign buffer. The samples were spun at 20°C at 20,000, 25,000 and 35,000 rpm for 24 hours to achieve equilibrium, as demonstrated by successive identical radial absorbance scans at 274 nm. The behavior of the peptide species at equilibrium is described by the following equation:

$$M_{buoy} = M_\omega (1 - \nu\rho)$$

where M_{buoy} is the measured buoyant weight, M_ω is the molecular weight in Daltons, ν is the partial specific volume of the sample and ρ is the density of the buffer solution. The partial specific volume of the sample and density of the buffer were calculated using SednTerp (ver. 1.06, University of New Hampshire) using the weighted average of the amino acid content. The samples were then spun at 20 °C at 20,000, 25,000 and 35,000 rpm for 24 hours to achieve equilibrium, and then successive radial absorbance scans were recorded. The peptide molecular weight and oligomerization behavior were determined by fitting the sedimentation equilibrium data from different initial loading concentrations and rotor speeds to various monomer-oligomer equilibrium schemes using WIN NonLIN (ver 1.035,

University of Connecticut) fitted to theoretical partial specific volume based on compositions (Johnson *et al.*, 1981). The partial specific volume of the sample and density of the buffer were calculated using SednTerp (ver. 1.06, University of New Hampshire) using the weighted average of the amino acid content. To obtain observed molecular weight and oligomerization states, absorbance results are converted to linear plots of ln(Absorbance) vs. r^2 (radius) to the theoretical monomer, dimer and trimer single-species models using following equation:

$$C_r = C_o \exp [M ((1-v_{bar}) \rho) \omega^2 (r^2 + r_o^2) / 2RT]$$

where C_r is the concentration at radius r (absorbance), C_o is the concentration at the meniscus, M is the monomer molecular weight, v_{bar} is the partical specific volume from SednTerp, ρ is the buffer density and ω is the angular velocity. Fig. 2-5 shows the residual and the linear plots of ln(Absorbance) vs. r^2 (radius) of a sample peptide.

F) Probe of secondary structure - circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy measures the differential absorption of circularly polarized light between different chiral molecules, thus serving as a structural probe of protein secondary structure. The absorption behavior of circularly polarized light obeys Beer's law ($A = \varepsilon Cl$), and is commonly expressed as molar ellipticity for biological measurements ($[\theta] = 3298\Delta\varepsilon$). Because CD is a low resolution technique, it provides information on the average secondary structure content of a peptide or protein sample. The typical CD spectra measures the far UV

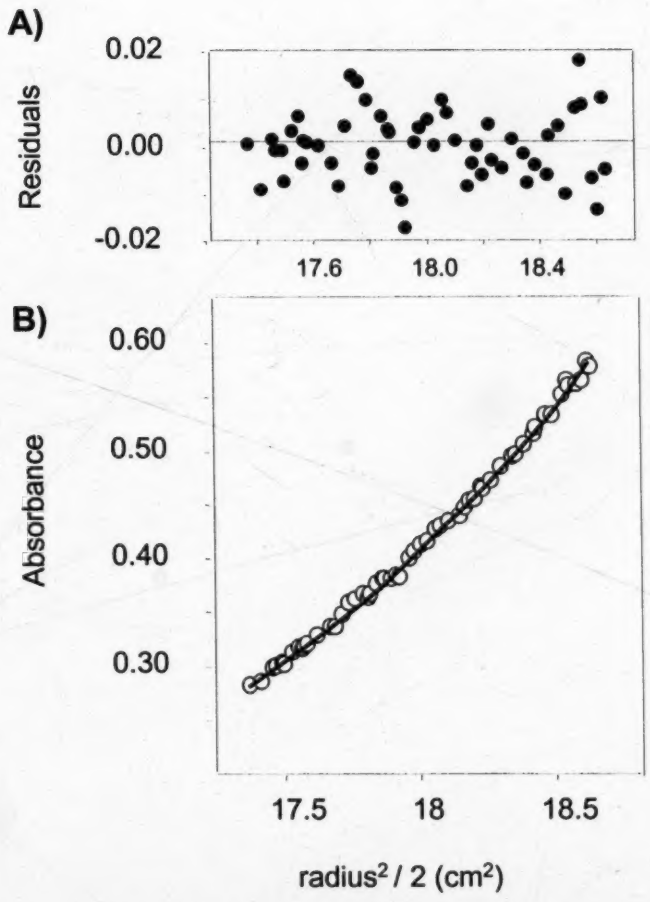


Fig. 2-5. Representative sedimentation equilibrium data **A)** Residual scatter showing the randomness and the unbiased fitting. **B)** Actual fit of the absorbance data versus radial distance squared divided by two.

spectrum of 190-250nm of peptides and proteins, which is very sensitive to protein secondary structure such as α-helix, β-sheet and random coil structure. For example, the α-helical conformation is characterized by double minima at 208 and 222 nm and the β-sheet conformation is characterized by a single minimum at 215 nm. Because of the simplicity of taking CD spectra, the linear deconvolution of any complex CD result of a given globular protein with multiple domains would provide an estimate of secondary structure presence and often served as the starting point for structural studies.

CD spectroscopy was performed on a Jasco-810 spectropolarimeter with constant N_2 flushing (Jasco Inc., Easton, MD.) for the estimate of secondary structure content and stability (Chen *et al.,* 1974). A Lauda circulating water bath was used to control the temperature of the optic cell chamber, where rectangular cells of 1 mm path length were used. The concentration of peptide stock solutions was determined by absorbance at 274 nm in 6 M Urea (extinction coefficient, ε = 1420 $cm^{-1} \cdot M^{-1}$, one tyrosine per peptide chain) and verified by amino acid analysis. For wavelength scan analysis, a 5 mg/ml stock solution of each peptide in 100 mM KCl, 50 mM PO_4, pH 7.0 (benign buffer) was diluted and scanned in the presence and absence of 50% trifluoroethanol (TFE). Mean residue molar ellipticity was calculated using the equation:

$$[\theta] = \theta_{obs} \text{ mrw } / 10\, l\, c$$

where θ_{obs} is the observed ellipticity in millidegrees, mrw is the mean residue molecular weight, l is the optical path length of the CD cell (cm), and c is the peptide

concentration (mg/mL). Each peptide spectrum was the average of eight wavelengths scans collected at 0.1 nm intervals from 195 to 250 nm. The uncertainty in the molar ellipticity values was +/-300 deg·cm^{-1}·dmol^{-1}.

i) Temperature-induced denaturation monitored by circular dichroism

Protein stability measurements were monitored at a wavelength of 222 nm, indicative of α-helical secondary structure. For thermal melting experiments, data points were taken at 1 °C intervals at a scan rate of 60 °C per hour. The temperature dependence of the mean residue ellipticity θ was fitted to obtain the fraction of the unfolded state, $f_{U(t)}$, using a non-linear least-squares algorithm assuming a two-state unfolding reaction with pre-transition (folded state, $\theta_{N(t)}$) and post-transition (unfolded state, $\theta_{U(t)}$) baseline corrections (Santoro & Bolen, 1988):

$$\theta_{(t)} = [(1 - f_{U(t)})\, \theta_{N(t)}] + [f_{U(t)}\, \theta_{U(t)}]$$

where the pre- and post-transition baselines corrections are assumed to be linearly-dependent on temperature. The assumption of the two-state unfolding transition of coiled-coils was recently discussed by Dragan and Privalov, 2002. In thermal denaturation experiments, coiled-coils showed small decreases in ellipticity prior to the large unfolding transition associated with coiled-coil unfolding. The small decreases were attributed to concentration-independent transitions due to end fraying and other molecular motion that does not lead to helix unfolding. These decreases are accounted for in our data set by the pre- and post-transition baselines. The

calculated fraction of the unfolded state, $f_{U(t)}$, is used to determine ΔG (Lavigne *et al.*, 1995):

$$f_{U(t)} = \exp [(-\Delta G_{app} / RT)] / [1 + \exp (-\Delta G_{app} / RT)],$$

where ΔG_{app} is the change in apparent Gibbs free energy of folding described by the Gibbs-Helmholtz equation:

$$\Delta G_{app} = \Delta H_{app} (1 - t / T_m) + \Delta Cp (t - T_m) - t \ln (t / T_m)$$

where t is the observed temperature, T_m is the temperature midpoint of the thermal transition, ΔH_{app} is the apparent enthalpy of unfolding and ΔCp is the change in heat capacity change associated with protein unfolding. ΔCp and ΔH are assumed to be constant within the temperature range (Makhatadze & Privalov, 1995; Gomez *et al.*, 1995). These thermodynamic parameters were fitted using the program Sigmaplot, and the derived fraction folded parameters were used to determine an apparent free energy stability term, $\Delta G_{app,X}$ using the following equation:

$$\Delta G_{app,X} = - RT \ln [f_{U(t)} / (1 - f_{U(t)})]$$

by plotting fraction folded data, then extrapolating to X°C.

ii) Chemical denaturation monitored by circular dichroism

For chemical denaturation experiments, the stock peptide solution was diluted with appropriate volumes of benign buffer and a stock solution of 10.0 M urea or 8.0 M GdnHCl in benign buffer to give a series of data points in increasing denaturant concentration. Data points were left to equilibrate overnight before

scanning and, to ensure accuracy, selected data points were rescanned to ascertain proper buffer equilibration. The data were fitted to a linear extraction method (LEM) described previously to determine denaturation midpoint, slope associated with the transition and the change in free energy associated with the transition, ΔG_u (Santoro & Bolen, 1988). A two-state unfolding model was used to derive peptide stability values from urea denaturation results:

$$f_f = ([\theta]_{obs}-[\theta]_u)/([\theta]_f-[\theta]_u)$$

where $[\theta]_f$ and $[\theta]_u$ represent the mean residue molar ellipticity for the fully folded and unfolded species, respectively, and the $[\theta]_{obs}$ is the observed molar ellipticity at a given denaturant concentration. The free energy of unfolding was derived from the equation:

$$\Delta G_U = RT \ln K_u$$

where K_u is the equilibrium constant of the unfolding process. In the case of disulfide-bridged peptides, where the unfolding process is concentration independent, K_u can be simplified as:

$$K_u = (1-f_f)/(f_f), \text{ thus}$$

$$\Delta G_U = RT \ln (1-f_f)/(f_f).$$

Estimates of the free energy of unfolding in the absence of denaturant, $\Delta G_{U(water)}$ and slope term m were obtained by extrapolation to zero (Pace,1986; Shortle, 1989):

$$\Delta G_U = \Delta G_{U(water)} - m[\text{denaturant}]$$

where m is the slope associated with unfolding.

G) Determination of enthalpy change and heat capacity by calorimetry

Besides thermal and chemical denaturation, reaction free energy and enthalpies can be measured directly by differential scanning calorimetry (DSC). Compared to thermal denaturation monitored by CD, this method has the advantage of more sensitive estimates of actual ΔH and ΔC_p because a CD temperature melting curve is generally insensitive to changes in ΔCp, and thus, can lead to an erroneous estimate of ΔH from multi-variables curve fitting. However, conventional calorimeter required large amounts of purified sample (0.5-2 mg) per experiment and demanded meticulous techniques to produce reliable results, thus limiting its application (Weber & Salemme, 2003). Nevertheless, newer technologies and the development of nano-scale instrumentation has facilitated the identification of intermediates during coiled-coil unfolding and revealed an example of leucine zipper unfolding that is more than just a simple two-state process (Dragan & Privalov, 2002).

Excess heat versus temperature for the peptides was determined using a Microcal differential scanning calorimeter (Microcal, Northampton, MA). The typical sample concentrations ranged from 105 to 140 μM of coiled-coil dimer, and peptides were dissolved in 100 mM KCl, 50 mM PO_4, pH 7.0 buffer. The sample solutions and buffer were filtered and degassed under vacuum and stored at 5 °C. Buffer scans were repeated until identical baselines were achieved. The heating rate was 60 °C per hour and the cooling rate was 90 °C per hour, with the excess heat

monitored from 5 to 70 °C. Each sample was heated and cooled for three cycles to ensure folding reversibility. Data analyses were carried out in Microcal Origin software (Microcal DSC, version 1.2a) using a two-state model with change in heat capacity.

Chapter III

Structural cassette mutagenesis in a *de novo* designed protein: proof of a novel concept for examining protein folding and stability

Stanley C. Kwok, Brian Tripet, Jeff H. Man, Mundeep S. Chana,
Pierre Lavigne, Colin T. Mant and Robert S. Hodges

A version of this chapter has been published in Biopolymers Peptide Science (1998) 47, 101-123

I) Introduction

An important goal in *de novo* design of proteins is to create novel proteins with defined structural folds mediating specific functions. However, before *de novo* design of complex proteins with new biological activities or binding properties can be achieved, a quantitative understanding of the relative contributions that all of the non-covalent interactions make in controlling protein conformation, folding, and stability is required. The difficulties of such an undertaking cannot be underestimated since, although nature generally employs only the two secondary structural motifs (α-helix and β-pleated sheet) as building blocks to assemble complex tertiary and quaternary structure, hundreds of variations of structure exist and no two proteins fold identically. In addition, the 20 naturally occurring amino

acids have unique secondary structure conformational preferences that have been experimentally demonstrated to influence the stability of a defined structural fold (O'Neil & Degrado, 1990; Lyu *et al.*, 1990; Serrano *et al.*, 1992; Horovitz *et al.*, 1992; Park *et al.*, 1993; Blaber *et al.*, 1994; Chakrabartty *et al.*, 1994; Creamer & Rose, 1994; Minor & Kim, 1994; Smith *et al.*, 1994; Zhou *et al.*, 1994a; Monera *et al.*, 1995). Recent research using chimeric small protein molecules has also shown that not all residues play an equally important role in a protein towards the folding and stability of that protein (Minor & Kim, 1996; Dalal *et al.*, 1997; Eidsness *et al.*, 1997). Complicating the matter still further is that the tendency of a stretch of amino acid residues to adopt a particular secondary structure is also context dependent, with the stabilizing or destabilizing contribution of an amino acid side chain toward a particular secondary structure being influenced by its environment, including a combination of short-range and long-range interactions with neighboring residues (Minor & Kim, 1994; Kuroda *et al.*, 1996; Li *et al.*, 1996; Yang *et al.*, 1997). Indeed, it was demonstrated that short stretches of peptides may adopt different secondary structures depending on the environment in which they are placed (Minor & Kim, 1996). Thus, the final conformation of a specific stretch of residues is determined not only by nearest-neighbor influences but also by the microenvironment effected by residues brought close to the polypeptide segment under consideration through folding of the entire polypeptide chain, i.e., stabilization of a particular secondary conformation by the intact tertiary structure of the protein itself. Indeed, an excellent example of such stabilization was reported by Gasset *et*

al., 1992 who found that isolated α-helical peptides of hamster prion form β-sheets, strongly indicating that the helical conformation required the intact tertiary structure of the protein. Such an observation also highlights a wider importance concerning secondary structure adaptation to its environment in that such conformational switching has recently been implicated in prion diseases. Pan *et al.* 1993 reported that the α-helical content of the normal cellular prion protein (PrP^{c}) diminishes and the amount of β-sheet increases in the abnormal disease-causing isoform (PrP^{Sc}). Indeed, Prusiner, 1997 pointed out that understanding how PrP^{c} unfolds and refolds into PrP^{Sc} may not only aid the prevention of prion diseases but may open new approaches to developing therapies for neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and amyotroresiphic lateral sclerosis.

Clearly, the solution to the protein folding problem lies in identifying stretches of residues crucial for the folding as well as defining the relative energetic contributions of short- and long-range interactions in these sequences (Munoz *et al.*, 1996; Pace *et al.*, 1996) The innate complexities of a global approach to such a solution, e.g., attempting to analyze the consequences of manipulating sequences of whole protein molecules through point mutation of individual residues and three-dimensional structure determination, an immensely time-consuming undertaking, points to the value of a minimalist option for such an investigation, i.e., selecting a well-defined model system that contains only one type of secondary structure (Nedwidek & Hecht, 1997). Such an approach has been extensively documented by our laboratory in its promotion of the two-stranded α-helical coiled coil motif as an

ideal model system to use in understanding the fundamentals of protein folding and stability and in testing the principles of *de novo* design (Hodges *et al.*, 1981; Lau *et al.*, 1984a,b; Hodges *et al.*, 1988; Hodges *et al.*, 1990; Hanson *et al.*, 1992; Hodges, 1992, 1996; Zhu *et al.*, 1992; Zhou *et al.*, 1992a,b; Monera *et al.*, 1993, 1994a,b, 1996a,b; Adamson *et al.*, 1993; Zhou *et al.*, 1993, 1994a,b; Zhu *et al.*, 1993; Su *et al.*, 1994; Kohn *et al.*, 1995a,b, 1997a,b, 1998a,b,c; Lavigne *et al.*, 1995; Yu *et al.*, 1996; Tripet *et al.*, 1996; Chao *et al.*, 1996; Houston *et al.*, 1996a,b). Since only one type of secondary structure, that of the easily monitored (by CD spectroscopy) α-helix, is present in such a model system, i.e., there are no β-turns, β-sheet, or undefined structure present, the interpretation of results is far less complicated. Intrinsic α-helical and β-sheet propensities of amino acid side chains have been demonstrated to be significant in stabilizing a defined secondary structure fold (Chakrabartty *et al.*, 1994; Minor *et al.*, 1994; Zhou *et al.*, 1994a; Monera *et al.*, 1995; Zhang *et al.*, 1992). In addition, the periodicity of polar and nonpolar residues along a sequence specifies whether a given sequence is consistent with α-helical or β-strand conformation, specifically amphipathic α-helix or β-strand (Xiong *et al.*, 1995), e.g., the 3-4 or 4-3 hydrophobic repeat characteristic of the amphipathic α-helices making up a two-stranded α-helical coiled coil or the alternating hydrophobic pattern characteristic of β-sheet. Indeed, armed with extensive evidence that the coiled coil motif is an ideal model for examining the rules of protein folding that direct polypeptide chain orientation and the oligomeric state, we wished to extend

further our understanding of such rules by addressing whether an amino acid sequence (or "cassette") with a defined secondary structure in a native protein can adopt a different conformation when placed into a different environment. Thus, we have designed a 39-residue chimeric peptide that forms a two-stranded α-helical coiled coil model and where an 11-residue cassette could be inserted into the center of the peptide between two nucleating sequences with defined secondary structure. In the present study to test the feasibility of the concept, a cassette from a β-sheet element of an immunoglobin fold was inserted between the two nucleating α-helices of the coiled coil. This β-strand sequence was chosen not only due to its high β-sheet propensity but also due to its reasonable α-helical propensity. In addition, while the selected sequence contained residues contributing to the hydrophobic core in stabilizing this strand in the β-sheet structure of the native fold, this sequence would, if it folded as a helix, contribute a hydrophobic core to stabilize the coiled coil structure, the intent being, as noted by Xiong *et al.,* 1995 that although amino acid residues may differ in their intrinsic preferences for one secondary structure over another, such preferences may perhaps be overwhelmed by the drive to form amphipathic structures capable of burying hydrophobic surface areas. By systematically enhancing the β-sheet propensity of this cassette region, we were able to examine whether such enhancement would change the overall secondary structure or modulate the stability of the model chimeric peptide. These results have clearly

demonstrated the feasibility of this model system to study short-range and long-range interactions in the stabilization of secondary structural folds.

II. Results and discussion

The major barrier to a full understanding of or merely designing an approach to understanding, why a particular residue sequence folds into a specific three-dimensional tertiary structure is the plethora of factors influencing protein stability and conformation. In addition to the well-known contributions of non-covalent interactions such as hydrogen bonding, van der Waals contacts, ionic (interchain and intrachain) interactions, and association of non-polar residues, must be added the potential of factors such as hydrophobic residue packing effects (Zhu *et al.,* 1993; Behe *et al.,* 1991; Alber, 1992; Harbury 1993, 1994), helix capping (Serrano *et al.,* 1992; Presta & Rose, 1988; Zhou *et al.,* 1994d; Tidor, 1994; Doig & Baldwin, 1995; Rohl *et al.,* 1996; Reymond *et al.,* 1997a,b), helix dipole effects (Kohn *et al.,* 1997c; Tidor, 1994; Fairman *et al.,* 1989; Aqvist *et al.,* 1991; Lockhart & Kim, 1993), and the above mentioned significant influence of intrinsic α-helix or β-sheet propensity of residues as well as the effects of hydrophobe periodicity. Other, more occasional, effects such as intrachain $i \rightarrow i + 3$ or $i \rightarrow i + 4$ nonpolar interactions along an α-helix (Padmanabhan, 1994) can also complicate the problem further.

Clearly, a minimalist approach in attempting to unravel the relative contributions of such a myriad of factors to protein structure is an appropriate starting point, i.e., isolate and analyze one factor's contribution at a time. Various

researchers have attempted such a minimalist approach by examining the structural capability of a region of sequence in the absence of the rest of the protein (Hamada *et al.*, 1995; Ramirez-Alvarado *et al.*, 1997; Reymond *et al.*, 1997b); such segments may already be known (or predicted) to form a particular secondary structure in the intact protein and their ability to fold into such structures (structure inducibility) may be examined in various media. Despite some interesting data arising from such work, the absence of the overall protein context makes for a limited extrapolation in light of the absence of the remaining part of the protein.

We believe our minimalist structural cassette approach, based on a *de novo* designed α-helical coiled coil motif, offers the potential for such an extrapolation. Although the advantages of such an approach are detailed below, including the aforementioned requirement to overcome concerns of protein context, it is immediately worth noting that such an approach offers a twofold advantage; thus, first, when the cassette is being inserted into the center of our *de novo* designed model protein, not only can we determine whether the cassette holder influences the cassette structure, but also how the cassette structure influences the cassette holder. Second, mutations in the cassette can examine short-range interactions. Thus, both short-range and long-range influences on protein structure may be analyzed.

A) Structural cassette mutagenesis (SCM): approach and design

Parameters relating intrinsic α-helical and β-strand propensity of naturally occurring amino acids have been extensively generated from both theoretical and

experimental approaches. Theoretically derived parameters include those obtained from statistical analysis (Chou & Fasman, 1974a,b; Chou & Fasman, 1978; Misra & Wong, 1997; Rost & Sander, 1994), as well as structural-based computer dynamic simulations (Creamer & Rose, 1992; Hermans *et al.,* 1992) During the past decade, many research groups have used defined model systems such as "host-guest" analysis of model peptides (O'Neil & Degrado, 1990; Lyu *et al.,* 1990; Park *et al.,* 1993; Blaber *et al.,* 1994; Chakrabartty *et al.,* 1994; Zhou *et al.,* 1994a, Monera *et al.,* 1995; Yang *et al.,* 1997; Wojcik *et al.,* 1990; Padmanabhan *et al.,* 1990; Merutka & Stellwagen, 1990a, Merutka *et al.,* 1990b) and site-directed mutagenesis of model proteins (Horovitz *et al.,* 1992; Minor & Kim, 1994; Smith *et al.,* 1994; Blaber *et al.,* 1993) as the stable host models to study α-helical or β-sheet propensity. Unfortunately, although many of these scales yield similar conformation preferences between amino acids, they are by no means identical; such discrepancies arise from context-dependent effects that are a feature of the specific experimental model employed. Thus, meaningful extrapolation of even the most rigorously developed intrinsic secondary structure propensities to the prediction of protein folding has remained elusive due to the lack of clear understanding on how both short-range and long-range interactions influence context-dependent protein structure and stability.

As described above, it is known that a sequence of amino acids that adopt a particular secondary structure can adopt a different conformation when placed in a different environment. Such a scenario suggested a promising step-by-step approach to delineating how the conformations of short peptide sequences are influenced by

their environment, an approach we have termed structural cassette mutagenesis (SCM).

B) SCM: Initial objective

The two-stranded α-helical coiled coil is now recognized as one of nature's favorite ways of creating a dimerization motif (Hodges, 1996). This structure consists of two amphipathic right-handed α-helices that adopt a left-handed super coil. To describe briefly: coiled coils can be considered as a repeating heptad of seven amino acid residues (***abcdefg***), where positions ***a*** and ***d*** are occupied by hydrophobic residues, this pattern being referred to as a 3-4 (or 4-3) hydrophobic repeat (Hodges *et al.*, 1981; Lau *et al.*, 1984a,b; Hodges *et al.*, 1988; Hodges *et al.*, 1990; Hanson *et al.*, 1992; Hodges, 1992, 1996; Zhu *et al.*, 1992; Zhou *et al.*, 1992a,b; Monera *et al.*, 1993, 1994a,b, 1996a,b; Adamson *et al.*, 1993; Zhou *et al.*, 1993, 1994a,b; Zhu *et al.*, 1993; Su *et al.*, 1994; Kohn *et al.*, 1995a,b, 1997a,b, 1998a,b,c; Lavigne *et al.*, 1995; Yu *et al.*, 1996; Tripet *et al.*, 1996; Chao *et al.*, 1996; Houston *et al.*, 1996a,b). In this initial study, we wished to design a two-stranded coiled coil with appropriate stability and length that can accept an 11-residue cassette representing structural elements of defined conformation from proteins whose three-dimensional structures have been determined. The cassette would represent β-strands, β-turns, loops, regions of undefined structure, and α-helical segments. The concept is to substitute such cassettes into the center of each polypeptide chain making up the two-stranded α-helical coiled coil, the coiled coil

thus providing a stable α-helical structure (nucleating α-helices) on each side of the cassette.

Questions to be addressed in this initial study were **1)** whether the coiled coil structure (the cassette holder) would nucleate the cassette into an α-helical fold regardless of the structure that the peptide sequence had in the original native protein, **2)** determining what properties the cassette would have to possess to prevent the cassette from adopting the α-helical fold, and **3)** what properties the cassette would have to possess to disrupt totally the coiled coil structure. Properties of the cassette that would be examined in addition to its known three-dimensional structure are the β-sheet, β-turn, and α-helical propensities of amino acids making up the cassette sequence, the amphipathicity of the cassette when in an a-helical conformation, and the hydrophobicity of the 3-4 repeat positions that make up the hydrophobic core of the cassette when in the two-stranded α-helical coiled coil conformation.

C) Propensity models

An important aspect in the development of our structural cassette model was to choose secondary structure propensity values that were intrinsic to each amino acid, i.e., such values should be as context independent as possible. Thus, the secondary structure propensity scoring carried out in this study was based on propensity parameters derived from two independent polypeptide models, both

designed to minimize external influences on the residue under consideration. Fig. 3-1A shows the helical net presentation of the model amphipathic α-helix used to determine α-helical propensities. The design was based on a single-stranded amphipathic α-helix with amino acid substitutions for propensity determinations being centrally located in the solvent-exposed non-polar face (Zhou *et al.,* 1994a) The uniform environment of only Ala residues surrounding the substitution site allowed the generation of a thermodynamic α-helical propensity scale devoid of unwanted intra- and inter-chain side-chain interactions. The model system used for the evaluation of β-sheet propensity (Fig. 3-1B), based on a mixed alpha/predominantly beta fold, shared a similar intent to the above amphipathic α-helical model in attempting to maximize the intrinsic nature of the generated parameters (Minor & Kim, 1994). Thus, substitutions of the 20 amino acids were made at a solvent-exposed site in the central position of a β-strand, where nearest neighbours were only Ala or Ser residues. The small side chains of these residues minimizes side-chain interactions with the residue at the substitution site.

The α-helix and β-sheet propensities of the 20 amino acids generated from the models of Zhou *et al.,* 1994a and Minor & Kim, 1994, respectively, are shown in Table 3-1. Specific points concerning the comparative α-helix and β-sheet propensities of amino acids will be made as later discussion requires. However, it is worth noting that the extremely high, in fact the highest, α-helix propensity of Ala (0.96), coupled with very low β-sheet propensity (0.0), was reflected in the presence

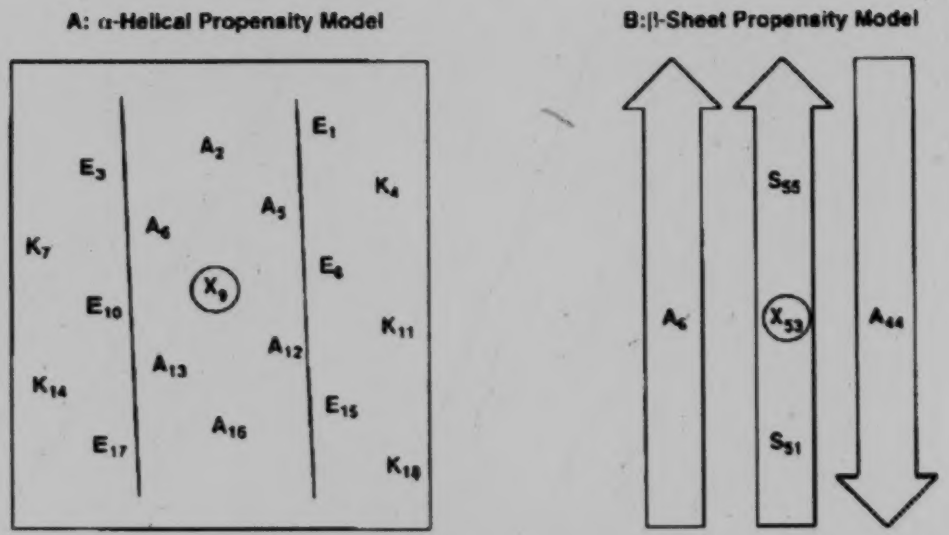


Fig. 3-1. Model peptide and protein designs from which were generated secondary structure propensity parameters used in this study. A: helical net representation of the model amphipathic α-helical peptide used by (Zhou et al., 1993b) to derive a-helical propensities. B: ribbon representation of the exposed β-strand of the model protein used by Minor and Kim (1994b) to derive β-sheet propensities. The site of amino acid substitution for all 20 amino acids is circled on both models and denoted by X.

Table 3-1 α-Helix and β-sheet Propensities of the Twenty Naturally Occurring Amino acids

α-helix propensity		β-sheet propensity	
Amino Acid	-ΔΔG (kcal/mol)	Amino Acid	-ΔΔG (kcal/mol)
Ala	0.96	Thr	1.10
Arg	0.90	Ile	1.00
Leu	0.81	Tyr	0.96
Lys	0.70	Phe	0.86
Met	0.67	Val	0.82
Gln	0.61	Met	0.72
Ile	0.59	Ser	0.70
Trp	0.49	Trp	0.54
Phe	0.48	Cys	0.52
Tyr	0.43	Leu	0.51
Cys	0.43	Arg	0.45
Val	0.42	Lys	0.27
Asn	0.33	Gln	0.23
Ser	0.33	Glu	0.01
His	0.33	Ala	0.0
Glu	0.32	His	-0.02
Thr	0.28	Asn	-0.08
Asp	0.21	Asp	-0.94
Gly	0.0	Gly	-1.20
Pro	-0.83	Pro	<-3.0

of only two Ala residues, instead of a maximum of four, surrounding the substitution site in the β-sheet model of Minor & Kim 1994; it was found that four Ala residues disrupted the β-sheet structure too much to use it as a model for propensity determinations.

D) Design of cassette holder coiled-coil

Based on the principles of *de novo* design of coiled coils (Hodges *et al.*, 1981; Lau *et al.*, 1984a,b; Hodges *et al.*, 1988; Hodges *et al.*, 1990; Hanson *et al.*, 1992; Hodges, 1992, 1996; Zhu *et al.*, 1992; Zhou *et al.*, 1992a,b; Monera *et al.*, 1993, 1994a,b, 1996a,b; Adamson *et al.*, 1993; Zhou *et al.*, 1993, 1994a,b; Zhu *et al.*, 1993; Su *et al.*, 1994; Kohn *et al.*, 1995a,b, 1997a,b, 1998a,b,c; Lavigne *et al.*, 1995; Yu *et al.*, 1996; Tripet *et al.*, 1996; Chao *et al.*, 1996; Houston *et al.*, 1996a), we created a cassette holder where the 3-4 hydrophobic repeat consisted of Val residues at position ***a*** and Leu residues at position ***d*** of the heptad repeat ***abcdefg***, these residues represent the core of the two-stranded α-helical coiled-coil.

From Fig. 3-2 (top), the N-terminal position of the cassette holder contains three hydrophobes (two Val residues and one Leu residue in the ***a*** and ***d*** positions) in the hydrophobic core plus the addition of an interchain disulphide bridge at the N-terminal of the sequence. This Cys-Gly-Gly sequence in each polypeptide chain allows a flexible linker and formation of the disulphide bridge to eliminate the monomer- dimer equilibrium, thus making guanidine hydrochloride denaturation

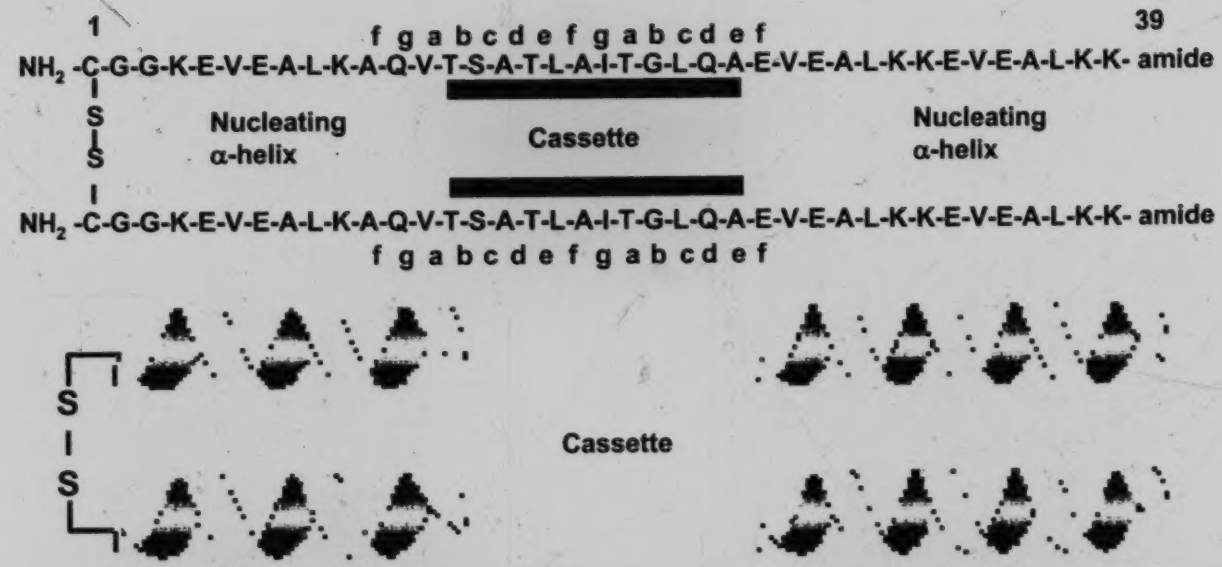


Figure 3-2 shows the amino acid sequence of the 78-residue disulfide bridge two-stranded α-helical coiled coil that is used as a cassette holder. The cassette holder consists of two nucleating α-helices at the N- and C-termini of each 39-residue peptide polypeptide chain. The hydrophobic residues at positions ***a*** and ***d*** of the heptad repeat denoted ***abcdefg*** that form the hydrophobic core between the two α-helices. The 11-residue cassette region in the center of the cassette holder is indicated by the solid bar. The hydrophobic residues of the cassette that are in positions ***a*** and ***d*** are circled and contribute to a continuous 3-4 hydrophobic repeat characteristic of a coiled-coil motif, along the length of the 39-residue polypeptide chain. The bottom section is a schematic of the cassette (denoted as an undefined conformation) and the cassette holder with two defined nucleating α-helices.

studies concentration independent (Zhou *et al.*, 1992a) In addition, the disulphide bond ensures the polypeptide chains of the coiled coil are parallel and in register; also, the disulphide bond would aid in the control of the oligomerization state maintaining a two-stranded coiled coil. The C-terminal region of the polypeptide chain contains two heptads with four hydrophobes (two Val residues and two Leu residues in the ***a*** and ***d*** positions) in the hydrophobic core. The N-terminal amino acid retains its free α-amino group to enhance peptide solubility.

Several other design features of the nucleating α-helices making up the cassette holder are also apparent from the sequence shown in Fig. 3-2 (top). Thus, positions ***g*** and ***e*** of the heptad repeat are occupied by Glu and Lys residues, respectively, allowing the potential for interchain salt bridge formation ($i \rightarrow i' + 5$ or ***g*** to ***e'***) at the coiled coil dimer interface, further strengthening the attraction between α-helices (Hodges, 1992; Monera *et al.*, 1994; Zhou *et al.*, 1994b,c; Yu *et al.*, 1996; Kohn *et al.*, 1997c); in addition, intrachain salt bridge formation between Lys residues at position ***e*** or ***f*** and Glu residues at position ***b*** ($i \rightarrow i + 3$ or $i \rightarrow i + 4$) enhances the α-helical nucleation potential (Monera *et al.*, 1994; Yu *et al.*, 1996; Kohn *et al.*, 1997b). The presence of Ala residues with very high α-helical propensity (Table 3-1) at position ***c*** and, frequently, at position ***f*** of the heptad repeat further assures the desired α-helical conformation for the model nucleating α-helices. Lys residues at the remainder of the ***f*** positions in the cassette holder (residues 4, 32, and 39) served to aid peptide solubility and avoid aggregation. Ala residues (positions 11***f*** and 25***f***) replaced Lys residues to ensure no intrachain

electrostatic interactions could occur between these positions and positions in the cassette. Glu was also replaced by Gln at position 12*g* to avoid any possible $i \rightarrow i' + 5$ electrostatic interactions with the cassette sequence.

E) Selection of cassette for insertion into coiled-coil

A fundamental requirement for the initial selection of the cassette region was a native peptide sequence that exhibited both α-helix and β-sheet propensity to a reasonable degree. Progressive mutation of this sequence to effect changes in the preference for one or the other secondary structure (in this case a preference for β-sheet) should then be experimentally feasible. In the present study, a sequence from a β-sheet element of an immunoglobulin fold appeared to serve our purpose.

The first immunoglobulin fold of human immunoglobulin G (IgG) fragment "Fab New" consists of 103 residues that fold into an all β-sheet tertiary fold (Fig. 3-3). However, a structural prediction strategy utilizing secondary structure propensities (Table 3-1) revealed a short stretch of the Fab New sequence that exhibited both the desired β-sheet potential and α-helical propensity (Fig. 3-4). Thus, Fig. 3-4 illustrates the α-helix (A) and β-sheet (B) propensity of residues 1→108 of the Fab New fragment (note that the black bars in panel B denote confirmed β-strand segments in the native protein). It is clear from Fig. 3-4 that the 11-residue sequence denoted by the grey bars (residues 64-74) is a β-strand in the X-ray structure (Fig. 3-3) and represents the longest continuous stretch of amino acids

Fig. 3-3. Ribbon drawings of an immunoglobulin fold (residues 1–103) of human immunoglobulin λ light chain of "Fab New" denoted 7FABL in the Brookhaven Protein Database. The top panel shows the immunoglobulin fold with the β-strand (colored white) that was used as a cassette inserted into the two-stranded α-helical coiled coil cassette holder. The bottom panel shows a side view of the immunoglobulin fold indicating the four hydrophobic side chains (colored white) , Ala 66, Leu 68, Ile 70, and Leu 73, buried in the hydrophobic core and the five hydrophilic side chains (Thr 64, Ser 65, Thr 67, Thr 71, and Gln 74) and one small hydrophobic side chain (Ala 69; colored green) , all protruding into the solvent. This 11-residue b-strand (64–74) is referred to as the native sequence cassette as shown in Fig. 3-2 (residues 14–24).

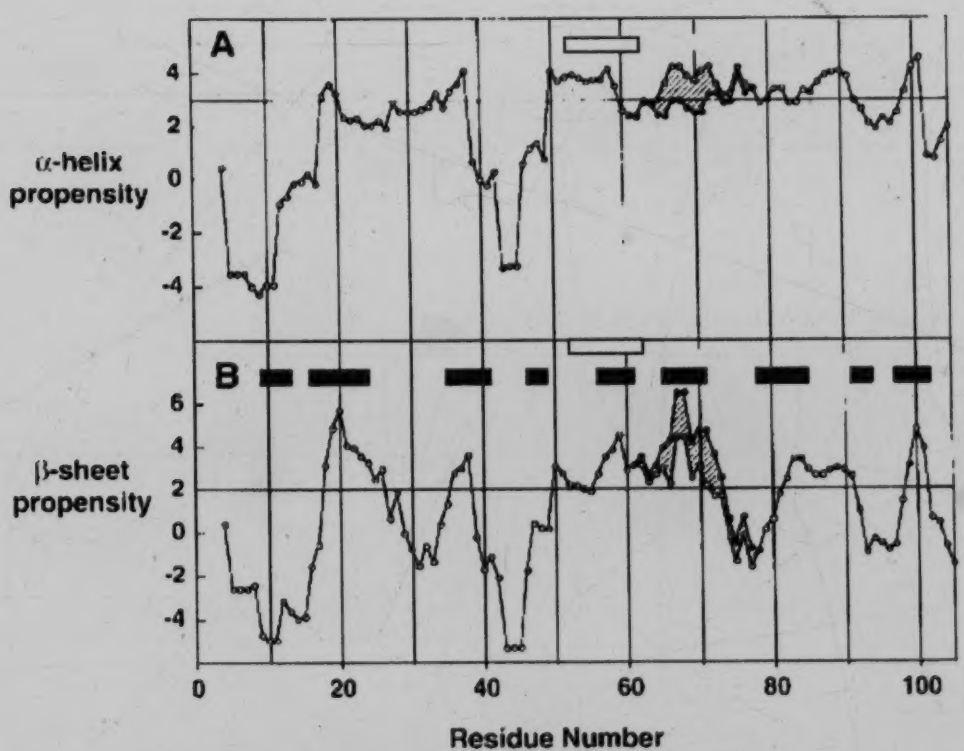


Fig. 3-4. The α-helix and β-sheet propensity scoring of Fab New (residues 1–108) . Each data point is the sum of seven consecutive secondary structure propensity scores (see Table 3-1) from the N-terminal of the protein. For example, the sum of residues 1–7 is plotted at position 4, and the sum of residues 2–8 is plotted at position 5, and so on. Panel A: plot of a-helical propensity scoring with a calculated cutoff value of 2.8, based on a consensus cutoff value from scoring of other a-helical proteins. Panel B: plot of b-sheet propensity with a predetermined cutoff value of 2.0, thus showing regions of high b-sheet propensity. The grey bars in Panels A and B represent residues 64–74 of Fab New, which were inserted as the cassette into the a-helical coiled coil model; the black bars in Panel B represent stretches of residues in Fab New that are known to adopt b-strand conformation from the third-dimensional x-ray structure. The white points denote scoring from the native Fab New sequence. The black points denote scoring from a mutation of this Fab New sequence, where the secondary structural propensity of the native cassette region has been altered by substitution with four Thr residues; the peptide sequence corresponding to this altered native region is denoted T4b in Fig. 3-7. The hatched areas denote the change in propensity scoring for these mutations.

with the potential for forming either α-helical or β-sheet structure based on prediction. Hence, this sequence was chosen as the "native" cassette (residues 14-24 in Fig. 3-2) to be inserted between the nucleating α-helices. Indeed, this 11-residue segment has the added advantage of being a good compromise for the average length of secondary structure elements found in proteins, being a little bigger than the average length of β-strands but representative of the length of α-helical segments (Chou & Fasman, 1974a,b, 1978; Chen *et al.*, 1974).

From Fig. 3-3 (top) the β-strand colored white represents this chosen cassette sequence, with the bottom panel indicating the four hydrophobic side chains (colored white) , Ala 66, Leu 68, Ile 70, and Leu 73 (Ala 16, Leu 18, Ile 20, and Leu 23, respectively, in Fig. 3-2). Note that these four hydrophobes are buried in the hydrophobic core of the immunoglobulin fold, while the hydrophilic side chains (colored green) Thr 64, Ser 65, Thr 67, Thr 71, and Gln 74 (Thr 14, Ser 15, Thr 17, Thr 21, and Gln 24, respectively, in Fig. 3-2) and one small hydrophobic side chain (also colored green), Ala 69 (Ala 19 in Fig. 3-2), all protrude into the solvent, highlighting the amphipathic nature of this β-strand.

Further advantages of the 11-residue native cassette become apparent when this sequence is illustrated as a helical coiled coil representation (Fig. 3-5). Thus, assuming a complete coiled coil conformation throughout the entire cassette and cassette holder sequence (Fig. 3-2), three of the hydrophobes in the cassette (Ala 16, Ile 20, Leu 23) would form a hydrophobic face continuous with the flanking α-helices, and hence contribute to the hydrophobic core of the coiled coil. The cassette

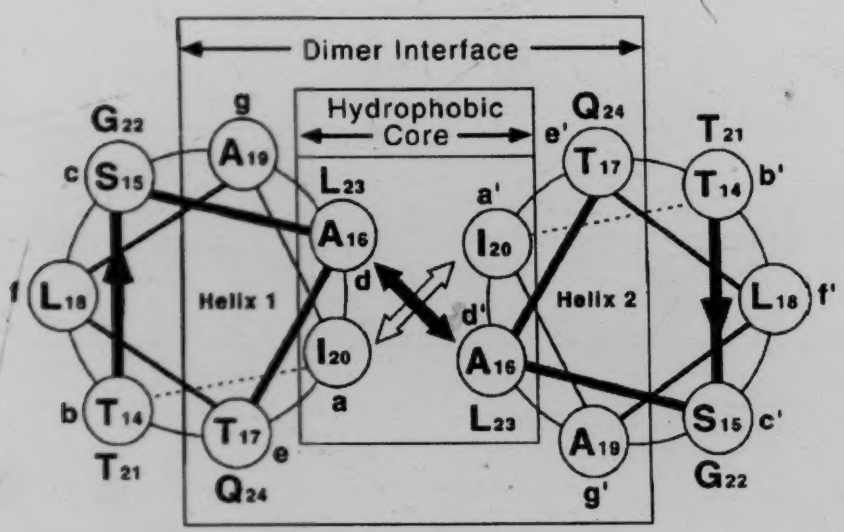


Fig. 3-5. Helical coiled coil representation of the 11-residue β-strand cassette denoted as residues 14–24 in the sequence, shown in Fig. 3-2. The view proceeds from T14 to Q24 into the plane of the paper. Position ***a*** and ***d*** and ***a′*** and ***d′*** of the heptad repeat ***abcdefg*** are denoted as making up the hydrophobic core; the side-chain to side-chain interactions (denoted by arrows) are between ***a*** and ***a′*** and ***d*** and ***d′***. Positions ***a, d, e, g,*** and ***a′, d′, e′, g′*** are denoted as making up the dimer interface.

would then represent approximately one-third of the coiled coil structure. This synchronicity of the β-strand sequence with the 3-4 hydrophobic repeat characteristic of the α-helical coiled coil structure made the immediate goal of transforming a β-chimera to α-helical motif seem even more favourable.

Another point to note from Fig. 3-5 is the presence of Leu 18 in the center of the hydrophilic face of the amphipathic helix. The equivalent of this residue in the native immunoglobulin fold, Leu 68, was shown to be buried in the hydrophobic core of the β-sheet structure (Fig. 3-3); in contrast, when represented in our α-helical conformation (Fig. 3-5), this residue is protruding into the solvent. This is consistent with previous observations that, for a given peptide sequence designed to produce an amphipathic β-strand, a consequent β-sheet structure will bury more hydrophobes than is possible through a two-stranded α-helical coiled coil conformation.

Fig. 3-6 displays as a space-filling model of the proposed native β-strand cassette inserted as an α-helix between the two nucleating α-helices. The hydrophobic and hydrophilic faces of the resulting single-stranded amphipathic α-helices are shown in the left and middle models, respectively. Note that the nonpolar residues of the 3-4 repeat making up the hydrophobic face along the whole length of the helix (left) are shown in blue; in contrast, the polar residues of the cassette region only are shown in green (middle; note Leu 18, in blue, in the center of this hydrophilic face). The model of the resulting α-helical coiled coil formed by two interacting single-stranded α-helices is shown at the right. Note that the non-polar

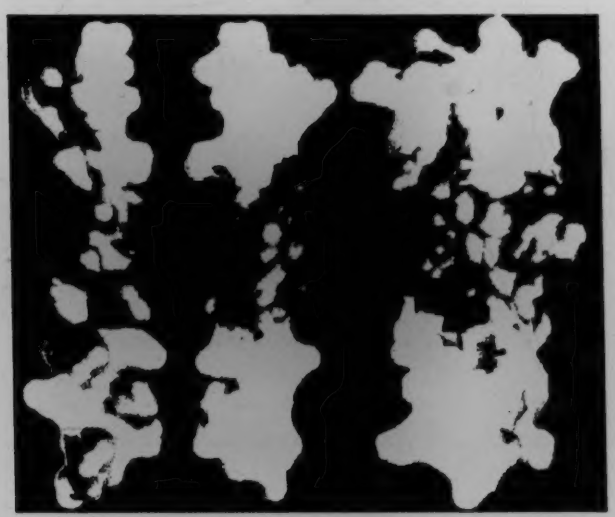


Fig. 3-6. Molecular surface model displaying proposed native β-strand cassette (sequence shown in Fig. 3-2) inserted as an α-helix between two nucleating α-helices (sequences shown in Fig. 3-2). At left is shown the hydrophobic face of the subsequent single-stranded amphipathic α-helix with the non-polar residues of the 3-4 heptad repeat shown in blue. In the middle is shown the hydrophilic face of the helix with the polar residues of the cassette region shown in green; the nonpolar residue (colored blue) in the center of this hydrophilic face is Leu 18 (see Fig. 3-2). At right is shown the a-helical coiled coil formed by two single-stranded α– helices interacting in an in-register and parallel orientation; the non-polar residues in the 3-4 heptad repeat and making up the hydrophobic core are buried at the dimer interface. Leu 18 (blue) is shown on the right and left side of the dimer.

residues in the 3-4 heptad repeat and making up the hydrophobic core are buried at the dimer interface. It can also be seen that Leu 18 in both strands of the dimer remains exposed on the surface of the coiled coil (Fig. 3-5 and 3-6).

F) Helicity and stability of the native model coiled-coils

Prior to investigating experimentally the suitability of the native protein as the starting point for subsequent structural mutagenesis, it was necessary to design a control protein as a reference, i.e., with a coiled coil stability higher than that of our native model, as well as any models containing mutant cassettes, since the majority of such cassettes would presumably be destabilizing to the molecule. The sequence of the control 11-residue cassette chosen (Fig. 3-7) was identical to that of the cassette holder sequence (residues 28-38; Fig. 3-2). Large hydrophobes are now in all requisite positions of the hydrophobic 3-4 repeat along the entire peptide sequence (note especially the replacement of the initially hydrophobic Ala side chain with that of the very hydrophobic Leu side chain at position 16***d***; fig 3-2 and 3-7) ; in addition, there is now potential for both interchain interactions ($i \rightarrow i' + 5$ interactions between Glu 19 → Lys 24) and intrachain interactions ($i \rightarrow i + 3$ between Lys 18 → Glu 21 and Glu 21 → Lys 24; $i \rightarrow i + 4$ interactions between Glu 14 → Lys 18, Lys 17 → Glu 21 and Lys 24 → Glu 28) involving residues in the control cassette sequence compared to the native sequence where, due to the absence of any potentially charged side chains, such interactions are not possible. Taking all

Peptide Name	Amino Acid Sequence of Cassettes
	1 14 24 39 c..... b c d e f g a b c d e .....f
Control	H_2N-..... E A L K K E V E A L K .....-amide
Native	H_2N-..... T S A T L A I T G L Q .....-amide
T1 (24)	H_2N-..... T S A T L A I T A L T .....-amide
T2a (22,24)	H_2N-..... T S A T L A I T T L T .....-amide
T2b (19,24)	H_2N-..... T S A T L T I T G L T .....-amide
T3a (19,22,24)	H_2N-..... T S A T L T I T T L T .....-amide
T3b (18,19,24)	H_2N-..... T S A T T T I T G L T .....-amide
T4a (18,19,22,24)	H_2N-..... T S A T T T I T T L T .....-amide
T4b (15,18,19,24)	H_2N-..... T T A T T T I T G L T .....-amide
T5a (15,18,19,22,24)	H_2N-..... T T A T T T I T T L T .....-amide
I20A	H_2N-..... T S A T L A A T G L Q .....-amide
I20T	H_2N-..... T S A T L A T T G L Q .....-amide

Fig. 3-7. Amino acid sequences of cassettes that were inserted into the centre of the two-stranded α-helical coiled coil cassette holder. The "Control" sequence contains as its cassette region (residues 14–24) the same sequence (EALKKEVEALK) as residues 28–38 of the nucleating α-helices (Fig. 3-2) . The "Native" sequence is the β-sheet cassette sequence taken from the immunoglobulin fold (Fig. 3-3). Peptides denoted T1–T5 are analogues of the native sequence, containing 1–5 Thr residues are substituted in positions ***c, e, f,*** or ***g,*** such positions being outside of the hydrophobic core of the a-helical coiled coil (Figure 3-5) . Peptides I20A and I20T represent mutations of the native sequence, where Ala or Thr, respectively, have replaced Ile 20 in the hydrophobic core of the coiled coil (Fig. 3-5). Nomenclature: for example, T5a represents five Thr substitutions in the native sequence at positions 15, 18, 19, 22, and 24; the "b" series represents Thr substitutions in the native sequence, where Gly 22 is not substituted; the "a" series represents Thr substitutions in the native sequence, where Gly 22 is substituted by Thr.

of the above into account, it would be expected that a coiled-coil containing this control cassette inserted in the cassette holder would be significantly more stable than any coiled coil that might be formed with the insertion of the native sequence. Note also that the potential for any attractive inter-chain $i \rightarrow i' + 5$ interactions between a Glu residue at position 12***g*** of the N-terminal nucleating helix (Fig. 3-2) and the Lys residue at position 17***e*** in the control cassette sequence (Fig. 3-7) has been eliminated by the presence of a neutral Gln residue at position 12 of the nucleating helix; the substitution of Gln for Glu at this position of the cassette holder was originally designed for just such a contingency, i.e., to minimize short-range interactions between the cassette holder and inserted cassette sequences that would complicate unnecessarily interpretation of stability data.

From Fig. 3-8 (top left), showing the CD spectra of the control and native oxidized (ox; disulphide-bridged) two-stranded proteins under benign conditions, it is clear that both cassette sequences inserted into the cassette holder resulted in high helicity (data summarized in Table 3-2); indeed, the helicity values of both proteins are almost identical (Table 3-2) and their CD spectra are characteristic of two-stranded α-helical coiled coils – namely, double minima at 220 and 208 nm, with a 222/208 ratio of 1.03 (Yu *et al.,* 1996) Even the addition of 50% trifluoroethanol (TFE), a helix-inducing solvent (Lau *et al.,* 1984a,b; Cooper & Woody, 1990; Sonnichsen *et al.,* 1992), caused no increase in helicity for both the oxidized control and the native proteins at 4 °C in benign medium whether in the reduced or oxidized

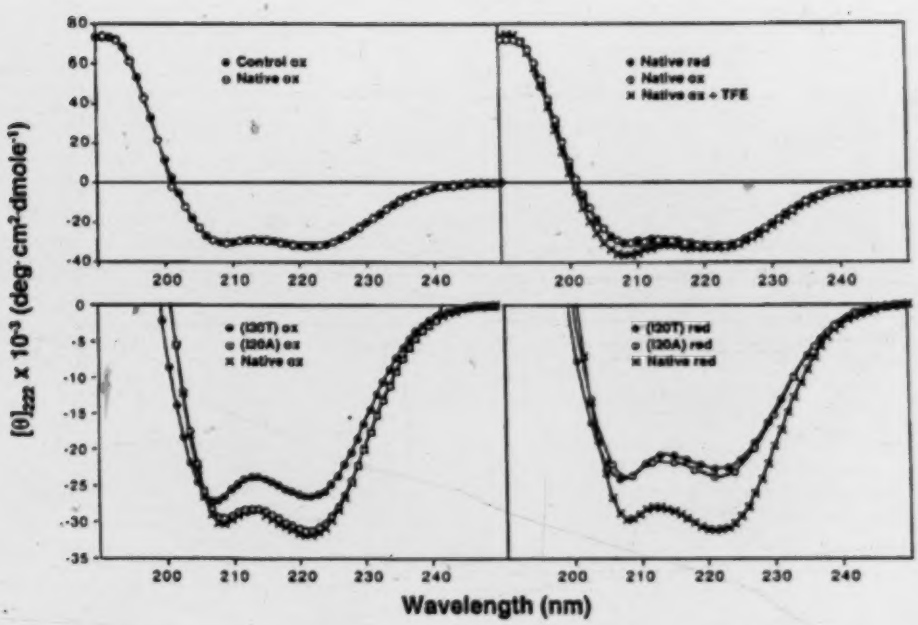


Fig. 3-8. CD spectra of the cassette holder (Fig. 3-2) containing various cassette sequences (Fig. 3-7) at 25°C in 50 m*M* KH_2PO_4 , 100 m*M* KCl, pH 7.0, in the absence (''benign'') or presence (/TFE) of 50% (v/v) TFE. The sequences of the cassettes (denoted Control, Native, I20T and I20A) inserted between the nucleating α-helices are shown in Fig. 3-7; ox (oxidized) and red (reduced) denote the presence and absence, respectively, of a disulphide bridge between the two interacting a-helical strands (see Fig. 3-2).

Table 3-2 Results and HPLC Data of the Oxidized and Reduced Coiled-Coils

	Oxidized[b]							Reduced[b]				
	$[\theta]_{222}$[c]							$[\theta]_{222}$				
Peptide Name[a]	Benign (25°C)	Benign (4°C)	50% TFE	% Helix[d]	$[Gdn \cdot HCl]_{1/2}$[e] (M)	m[f] (kcal · mol · M^{-1})	$\Delta\Delta G^u$[g] (kcal · mol)	Benign (25°C)	Benign (4°C)	50% TFE	% Helix	t_R[h]
Native	−31,800	−33,400	−33,200	90.8 (95.4)	3.2	1.82	—	−31,400	−33,100	−32,000	89.7 (94.6)	38.6
Control	−32,300	−35,900	−34,500	92.3 (102)	5.6	1.93	4.50	−32,000	−35,200	−34,500	91.4 (100)	42.2
T1 (24)	−30,900	−32,800	−32,100	88.3 (93.7)	2.9	1.68	−0.54	−30,700	−31,900	−32,800	88.8 (91.1)	39.2
T2a (22,24)	−32,100	—	−32,900	91.7	3.5	1.87	0.57	−31,100	—	−33,200	89.4	40.5
T2b (19,24)	−32,400	−33,800	−33,600	92.5 (96.6)	2.7	1.64	−0.89	−29,200	−32,100	−34,000	83.4 (91.7)	40.5
T3a (19,22,24)	−32,800	—	−34,600	93.7	3.3	1.53	0.17	−30,300	—	−34,600	86.6	37.1
T3b (18,19,24)	−31,400	−32,300	−32,900	89.7 (92.3)	2.4	1.70	−1.45	−27,800	−30,100	−31,900	79.4 (86.0)	37.2
T4a (18,19,22,24)	−31,700	—	−34,500	90.6	3.0	1.85	−0.38	−29,800	—	−34,800	85.1	38.7
T4b (15,18,19,24)	−30,900	−32,900	−32,700	88.3 (94.0)	2.2	1.69	−1.81	−24,000	−26,800	−32,800	68.6 (76.6)	38.2
T5a (15,18,19,22,24)	−31,300	—	−33,000	89.4	2.7	1.86	−0.95	−28,200	—	−32,900	80.6	38.7
I20A	−28,700	−32,000	−32,500	82.0 (91.4)	1.6	1.64	−2.86	−23,700	−30,300	−33,600	67.7 (86.6)	37.6
I20T	−26,300	−28,900	−32,700	75.1 (82.6)	1.4	1.71	−3.20	−23,200	−26,300	−33,800	66.3 (75.2)	36.8

a Peptide cassette sequences and nomenclature are shown in Fig. 3-7.

b Oxidized and Reduced peptides denote the presence or absence of a disulfide bridge. Peptide concentration ranges from 200 to 300 uM.

c $[\theta]_{222}$ is the mean residue molar ellipticity measured at 222 nm in a 50 mM $KHPO_4$, 100 mM KCl buffer, pH 7.0.

d % Helix was calculated from $[\theta]_{222}$ based on an ellipticity value of -36,000 for 100% α-helical content derived from the equation $X_H^n = X_H^\infty (1\text{-}k/n)$, where X_H^∞ is - 37,400, the wavelength dependent constant, k, is 2.5, and n is 36 residues (Chen *et al.*, 1974).

e $[Gdn{\cdot}HCl]_{1/2}$ is the denaturation midpoint of the two-state unfolding of an α-helical coiled-coil to a random coil. Guanidinium denaturation were carried out at 25 °C after overnight equilibration at room temperature. $[GdnHCl]_{1/2}$ is reproducible within +/- 0.05 M when the same stock solution of guanidinium hydrochloride was used. Values precised to 0.05 M were reported. Concentration of peptides ranges from 87 to 104uM.

f m is the slope term defined by the linear extrapolation equation $\Delta G_u = \Delta G_u(H_20) - m$ [denaturant], and $\Delta G_u(H_20)$ is the free energy of denaturation in the absence of denaturant, whereas ΔG_u is the free energy of unfolding at a given denaturant concentration. (See Chapter II)

g $\Delta\Delta G_u$, the change of the free energy of unfolding relative to peptide N14A,N15A,S23L , derived from the equation $\Delta\Delta G_u = ([denaturant])_{1/2,a} - [denaturant]_{1/2,b})(m_a + m_b)/2$, (Serrano & Fersht, 1992; see Chapter II).

states (Table 3-2). From Fig. 3-8 (top right), the shift of the absorbance at 208 nm relative to 222 nm is indicative of unraveling of the coiled coil, while maintaining the helicity of the two peptide strands, i.e., in the presence of the organic solvent, 50% TFE, tertiary and quaternary structure is lost (Lau *et al.,* 1984a,b; Cooper & Woody, 1990; Sonnichsen *et al.,* 1992) and the protein is in the form of two individual α-helices, dissociated but linked by a disulphide bridge. Reduction (red) of the disulphide bridge linking the two strands of the coiled coil also had a relatively negligible effect on the observed helicity of the control (Table 3-2) or native (Fig. 3-8, top right; Table 3-2) proteins, even in the presence of 50% TFE (Table 3-2), thus demonstrating the success of transforming a β-strand segment into an α-helix in our coiled coil model. Note that the formation of α-helix was favored despite the exposure of the hydrophobic residue Leu 18 on the outside of the cassette region of the coiled coil (Fig. 3-5), compared to its burial as part of a β-strand in the original IgG protein fold (Fig. 3-3). In addition, while the reduced proteins are remaining as coiled coils under benign conditions, they are present only as single helical strands in the presence of 50% TFE.

To verify the oligomerization state of our proteins, both size-exclusion chromatography and sedimentation equilibrium studies were used to show that the disulfide-bridged two-stranded α-helical coiled coils were monomeric in benign medium and that the reduced coiled coils were dimeric in benign medium.

Despite the similarity in the helical content of the control and native sequences, GdnHCl studies of the oxidized coiled-coils demonstrated a clear

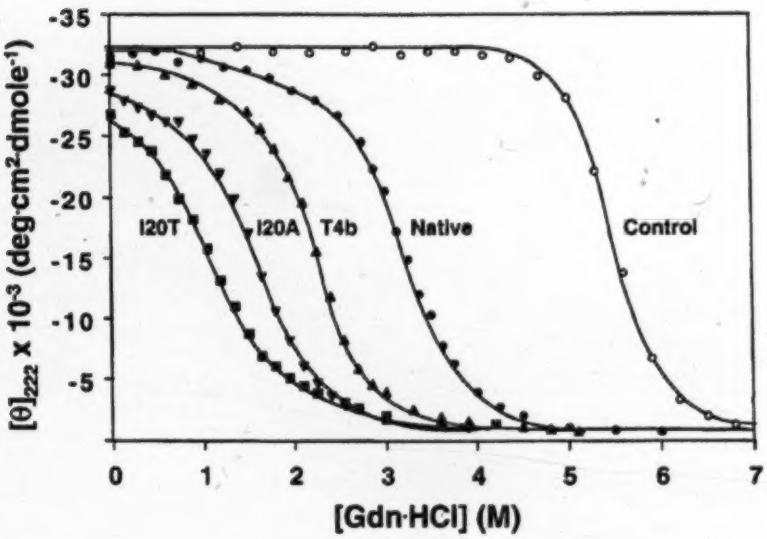


Fig. 3-9. Guanidine hydrochloride denaturation profiles of the cassette holder (Fig. 3-2) containing various cassette sequences (Fig. 3-7) . CD data were recorded at 25°C in 50m*M* KH_2PO_4 , 100m*M* KCl, pH 7.0, containing increasing concentrations of GdnHCl. The sequences of the cassettes (denoted Control, Native, T4b, I20T, and I20A) inserted between the nucleating α-helices are shown in Fig. 3-7.

difference in coiled coil stability (Fig. 3-9). Thus, the control sequence did indeed exhibit the high stability ($[GdnHCl]_{1/2}$ transition midpoint value of 5.6M; Table 3-2) required of a reference protein; in contrast, the native cassette sequence chosen to test the feasibility of the model destabilized the coiled coil, with an observed transition midpoint value of 3.2M (Table 3-2). It is important to note that, while the native cassette sequence (originally, of course, part of the β-sheet structure in an immunoglobulin fold) has destabilized the coiled coil structure quite considerably relative to the control cassette, it has not been destabilized to such an extent as to make it unsuitable for the purposes of our model. Indeed, as is clearly demonstrated later, the mid-range stability effected by the native cassette is ideal in allowing us to modify further the amino acid sequence in order to modulate coiled coil stability by varying the hydrophobicity and/or α-helix/β-sheet propensity of the cassette. In addition, the disulphide bond could be reduced to determine whether decreasing overall stability of the protein would allow the cassette region to disrupt the α-helical structure of the cassette holder.

Prior to discussing the effect on protein structure of subsequent modulation of the cassette structure, it is worthwhile to consider the successful transformation of β-strand to α-helix conformation in light of the relative contributions that different types of non-covalent interactions make to the secondary structure of the cassette sequence in both the native IgG protein as well as our *de novo* designed protein model. Specifically, it is important to be aware of any potentially unequal interactive factors, external to residue α-helical or β-sheet propensity, which may have a

bearing on the secondary structure preference of the native cassette sequence when removed from the native protein and inserted into a potential α-helical coiled coil motif. Thus, theoretically, the maximum number of hydrogen bonds that could be formed by all residues of the 11-residue cassette is the same (22), whether the sequence exists as a b-strand in the native IgG protein with interstrand hydrogen bonds stabilizing the β-sheet structure, or as an α-helix in the coiled coil motif, with intrachain hydrogen bonds stabilizing the individual α-helices. However, using the VADAR program, it was calculated from the X-ray crystal structure of the native IgG protein that only 18 out of a possible 22 bonds are formed by the native β-strand. Although the importance of hydrogen bond stabilization to secondary structure is still a controversial subject (Pace, 2001; Scholtz *et al.,* 1993), there might still have been some concern that the greater number of hydrogen bonds that would be formed by the 11-residue native sequence in an α-helical conformation might affect (i.e., help promote) the transformation from sheet to helix. However, it should be noted that, in general, the burial of hydrophobes is better achieved by an amphipathic β-sheet with alternating hydrophobic/hydrophilic residues (an "ideal" β-sheet), where every second residue can be buried, thus contributing significantly to protein stability; in contrast, an α-helix with hydrophobic residues in positions ***a*** and ***d*** (the 3-4 or 4-3 hydrophobic repeat discussed earlier) will only have 2 residues out of 7 buried in the hydrophobic core of a coiled coil, although residues in positions ***e*** and ***g*** that can interact with the hydrophobic core can add further stability (Hodges *et*

al., 1994). Indeed, the VADAR program calculated that the four hydrophobes (Ala 16, Leu 18, Ile 20, Leu 23; Fig. 3-2) in the native cassette sequence are 99.2% buried when this sequence is in the native IgG protein as a β-strand; in contrast, these hydrohobes are only 68% buried when the sequence is in an α-helical coiled coil conformation (or 85.5% buried if the surface-exposed Leu 18 is omitted from the calculation; Fig. 3-5 and 3-6). Thus, the greater potential for hydrophobic interactions in the β-sheet conformation compared to the coiled coil motif would be expected to offset any possible destabilization effects due to the lesser number of hydrogen bonds in the former compared to the latter, i.e., all things being equal and without any external (contextual) influences, it might be expected that the native cassette sequence would show a preference for β-sheet conformation. However, we have shown that the nucleating α-helices of our *de novo* designed protein are readily able to transform the β-strand to an α-helix, laying to rest any concerns about the inequality of hydrogen bond contributions between the two structures or whether our cassette holder could indeed overcome the more favorable hydrophobe-shielding aspect of the β-sheet structure compared to our α-helical coiled coil.

G) Effect of increasing β-sheet propensity of the cassette on overall coiled-coil stability

Having confirmed that a β-sheet segment (cassette) would fold as an α-helix when inserted into an α-helical protein (cassette holder), we now wished to determine whether substituting high β-sheet propensity amino acids in the

hydrophilic face of the cassette could switch the cassette from α-helix to β-sheet or an undefined structure or whether these changes would simply destabilize the coiled coil structure or have more dramatic effects and disrupt the α-helical structure of the cassette holder. From Table 3-1, threonine has very high β-sheet propensity but only low α-helical propensity, making this residue very suitable for our purposes. Thus, two series of cassette analogues were synthesized, with 1, 2, 3, 4, or 5 Thr residues being substituted at positions ***c, e, f,*** or ***g*** of the native cassette sequence, the β-sheet propensity increasing and the α-helix propensity decreasing with each successive Thr addition. From Table 3-2, one series of cassettes, denoted T1, T2b, T3b, and T4b for, respectively, 1, 2, 3, and 4 Thr substitutions, retained Gly 22; in contrast, the second series of cassettes, denoted T2a, T3a, T4a, and T5a for, respectively, 2, 3, 4, and 5 Thr substitutions, had Gly 22 replaced by one of the Thr residues. Thus, one cassette series (T1, T2b-T4b) allows insight into the effects on protein stability of substituting Thr, with very high β-sheet propensity, for residues of moderate to high a-helix propensity (Ser, Gln, Leu, Ala; Table 3-1) coupled with low to moderate β-sheet propensity save for the highly β-sheet disrupting Ala (Table 3-1); in contrast, the second cassette series (T2a-T5a) had the added dimension of substituting with Thr a residue (Gly), which provides flexibility and is known to disrupt both α-helix and β-sheet (Table 3-1).

From Table 3-2, CD spectral data indicate high helicity values for all of the oxidized (ox; disulphide-bridged) two-stranded proteins under benign conditions, the

values being comparable to those exhibited by both the native and control proteins. Thus, both series of high β-sheet propensity cassettes inserted into the cassette holder were still maintained in an α-helical conformation by the combined presence of the two nucleating α-helices and disulphide bridge. Also, in a similar manner to the control and native proteins, the addition of 50% TFE caused no significant increase in helicity of the oxidized coiled coil series examined (no Gly 22 substitution) from that observed at 4 °C in benign medium.

In contrast, a significant decrease of α-helical structure was apparent for some of the coiled coils in the absence of a disulphide bridge. From Table 3-2, the helicity of the proteins decreased with increasing numbers of high β-sheet propensity Thr substitutions in the inserted cassette, this decrease being clearly more evident with the cassette series where Gly 22 was retained (T1, 88.8% helix; T2b, 83.4%; T3b, 79.4%; T4b, 68.6%) compared to the series where this a-helix- and β-sheet-disrupting residue was substituted by Thr (T2a, 89.4% helix; T3a, 86.6%; T4a, 85.1%; T5a, 80.6%). Note that, even the two proteins with the lowest helicity at 25 °C (T3b and T4b) exhibited an increase in helicity at 4 °C, as expected, indicative of a shift from unfolded monomers to folded dimer at lower temperature and a decrease in the flexibility of the ends of the helices that are known to be frayed in coiled coils in solution (Zhu *et al.*, 1992; Zhou *et al.*, 1992a,b).

From Fig. 3-4, it can be seen that the structural propensity of the cassette denoted T4b has undergone a dramatic change with the substitution of four Thr residues into the native sequence. Thus, the β-sheet propensity scoring of this region

(Fig. 3-4B) has been significantly enhanced concomitant with a much reduced α-helical propensity scoring (Fig. 3-4A), in keeping with the high β-sheet propensity and low α-helix propensity of Thr (Table 3-1). In fact, the data in Fig. 3-4 would not predict α-helical structure in this region based on propensity data alone. Therefore, long-range interactions must be contributing substantially to the stability of the cassette. This is, of course, reflected by the decreased helicity of the peptides as the number of Thr residues increases in the cassette.

The decrease of α-helical structure by the progressive substitution of Thr residues into the cassette region seen in the reduced proteins (Table 3-2) is also paralleled by the stabilities of the oxidized coiled coils as determined by GdnHCl denaturation studies (Table 3-2). Thus, from Fig. 3-10 (top), the decrease in stability of the two series of oxidized proteins (illustrated in Fig. 3-9 for T4b relative to native) with increasing numbers of Thr substitutions in solvent-exposed regions is quite clear. Indeed, there was an essentially linear decrease in stability for both series of proteins (reflected also by the progressive decrease in ΔΔG values; Table 3-2), with the stability of the native coiled coil falling comfortably in line with the linear pattern of the cassette sequences retaining Gly 22 (T1, T2b, T3b, T4b). This observation is not surprising given that the native sequence, of course, also has Gly in this position. In contrast, although the protein series not containing Gly 22 also showed a linear decrease in stability (again, also reflected by decreasing ΔΔG values; Table 3-2), with an increasing number of Thr residues in the cassette sequence (T2a, T3a, T4a,

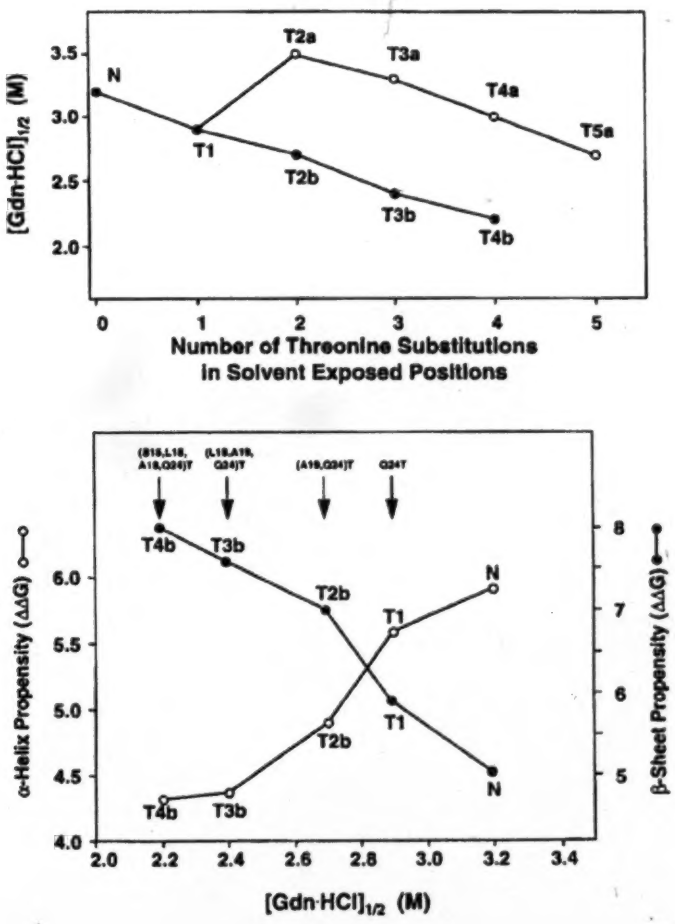


Fig. 3-10. Top: Effect on protein stability of cumulative Thr substitutions in the cassette region of the α-helical coiled coil. Bottom: relationship between a-helical or β-sheet propensity of cassette region and the stability of the α-helical coiled coils. The values for GdnHCl denaturation midpoint ($[GdnHCl]_{1/2}$) and differences in free energy of unfolding ($\Delta\Delta G$) were determined as described in Chapter II. Each point denoting secondary propensity scores (see Table 3-I) represents the sum of the propensity scores of all 11 residues in each mutant cassette. The sequences of the cassettes are shown in Fig. 3-7; above each mutant cassette denotion in the bottom panel (see arrows) is shown the specific residues in the cassette being substituted by Thr, e.g., for T2b, Thr is replacing both Ala 19 and Gln 24.

T5a), substitution of Thr for Gly (as noted above, a residue noted for its severe disruption of secondary structure propensity; Table 3-1) to produce the T2a cassette sequence increased considerably the stability of the resulting coiled coil relative to the native protein as well as the rest of the protein series retaining this residue; hence, the two-tiered stability profiles seen in Fig. 3-10 (top) for the oxidized proteins and decrease in helicity results shown in Table 3-2 for the reduced proteins.

Fig. 3-10 (bottom) offers an alternate illustration of the effects of cassette mutation on coiled coil stability. Thus, coiled coil stability for proteins with native T1, T2b, T3b, and T4b cassette insertions (as expressed by $[GdnHCl]_{1/2}$ values) has been linked with the overall α-helix or β-sheet propensities of the whole 11-residue cassette sequence, which are, of course, related to the number of Thr substitutions in the sequence. From Fig. 3-10 (bottom), the relationships of coiled coil stability with α-helix or β-sheet cassette propensities are shown quite vividly: a decrease in stability of the α-helical coiled coil with increasing β-sheet propensity of the cassette, i.e., an inverse relationship; conversely, an increase in stability of the a-helical coiled coil with increasing α-helix propensity of the cassette. In summary, although a switch in secondary structure from α-helix to β-sheet of the cassette region has yet to be achieved, likely due to the inability of the cassette to overcome the long-range interactions by the cassette holder stabilizing this region, the decrease in overall coiled coil stability has resulted simply from increasing the β-sheet propensity of a central minority sequence of the protein.

H) Effect of decreasing hydrophobicity of non-polar core of cassette on protein stability

It has been well documented that the stability of an α-helical coiled coil may be modulated by the nature of the side-chains at hydrophobic positions ***a*** and ***d*** of the heptad repeat ***abcdefg*** (Hodges *et al.*, 1990; Hanson *et al.*, 1992; Zhu *et al.*, 1992, 1993; Zhou *et al.*, 1992a,b,c, 1993; Adamson *et al.*, 1993). Thus, the less hydrophobic (or more hydrophilic) the side chain at one or more of these non polar positions, the less stable the resulting coiled coil. Indeed, even a single substitution in the hydrophobic core such as a mildly hydrophobic Ala residue for a very hydrophobic Leu residue at one site in the center heptad of a 35-residue (i.e., comparable in length to the present coiled coil model) two-stranded a-helical coiled coil has been shown to have a considerable destabilizing effect on the coiled coil structure (Hanson *et al.*, 1992; Zhu *et al.*, 1992; Zhou *et al.*, 1992a,b,c). Thus, the next logical step in assessing the feasibility of our mutagenesis approach was to modulate, i.e., decrease, the hydrophobicity of the cassette region in order to assess the effects of such a change on protein stability and/or conformation.

Initially, two cassette sequences with a single mutation (relative to the native cassette) in a hydrophobic core position have been synthesized. One sequence, denoted I20A (Fig. 3-7) has an Ile residue substituted by an Ala residue at position 20a; the second sequence, denoted I20T (Fig. 3-7), has an Ile residue substituted by a Thr residue at this position. Thus, a very hydrophobic residue with moderate α-

helical propensity and high β-sheet propensity (Ile; Table 3-1) has been replaced either by a moderately hydrophobic residue with high α-helical propensity and very low β-sheet propensity (Ala; Table 3-1) or by a polar, moderately hydrophilic residue with low α-helical propensity and high β-sheet propensity (Thr; Table 3-1). In other words, the large hydrophobic residue, Ile, is being substituted by Ala, a residue with the highest α-helical propensity or Thr, a residue with the highest β-sheet propensity.

From Table 3-2, it is apparent that a single substitution in the hydrophobic core of the cassette region had a destabilizing effect on protein α-helicity greater even than that observed for multiple substitutions with the high β-sheet propensity Thr residue in solvent-exposed positions. Thus, even in the presence of a disulphide bridge (oxidized), some loss in helicity was seen for both I20A (82% helicity at 25°C) and I20T (75%) compared to that of the native protein (90.8%); this is also illustrated by the CD spectra of these three proteins in Fig. 3-8 (bottom left). The helicity of both the mutant proteins is increased, as expected, at 4°C (91.4 and 82.6%, respectively, for I20A and I20T) compared to 25°C (82 and 75%, respectively, for I20A and I20T); note, however, that the relatively low helicity of I20T even at 4°C (the helicity of this protein only being maximized by the presence of 50% TFE; Table 3-2) underlines the particular destabilizing effect of the Ile → Thr substitution in the hydrophobic core. The results obtained with the reduced coiled coils were even more dramatic (Table 3-2 and Fig. 3-8, bottom right) in terms

of the loss of helicity compared to the native protein. Thus, at 25°C, the I20A and I20T proteins exhibited helicity values of only 67.7 and 66.3%, respectively, these values rising only to 86.6 and 75.2%, respectively, at 4°C.

The loss in α-helicity effected by the hydrophobic core mutations was also mirrored by the quite considerable loss in coiled coil stability (expressed as $[GdnHCl]_{1/2}$ values) of the I20A and I20T proteins compared to the native and, of course, control proteins, as clearly illustrated by the denaturation profiles shown in Fig. 3-9. From Table 3-2, the denaturation midpoint values for the oxidized I20A and I20T coiled coils have dropped to 1.6 and 1.4M, respectively, compared to 3.2 (native) and 5.6M (control). As noted above, the helicity loss resulting from these hydrophobic core mutations, the result of only one residue substitution, was comparable to that effected by multiple Thr substitutions in solvent exposed positions. In contrast, the loss in coiled coil stability resulting from these single hydrophobic core mutations actually exceeds that exhibited by even multiple Thr substitutions in the external (to the coiled coil) residue positions; witness, for example, the denaturation midpoint value (2.2M) of the T4b coiled coil (four Thr substitutions; Table 3-2, Fig. 3-9) compared to I20A (1.6M) and I20T (1.4M). Finally, it should be noted that the greater loss in α-helicity of the I20T protein compared to I20A, concomitant with the lower coiled coil stability of the former compared to the latter, can be attributed to the aforementioned properties of Thr and Ala; thus, it was not surprising that Ala, with much higher α-helical propensity and significantly greater hydrophobicity than Thr, would, when substituted into the

hydrophobic core of the cassette, result in a more stable molecule than that hosting a Thr insert in the same position.

The above comments concerning the link between side-chain hydrophobicity in the hydrophobic positions of the coiled coil and its subsequent stability now suggested another, high performance liquid chromatography based approach to relating these parameters. It is well documented that the nonpolar environment characteristic of RPC (hydrophobic stationary phase, nonpolar mobile phase) enhances α-helical conformation (in a manner similar to the addition of TFE) while disrupting tertiary and quaternary structure (Lau *et al.,* 1984; Zhou *et al.,* 1995). Thus, under reduced conditions, i.e., in the absence of a disulphide bridge, all of the peptides in the present study will be eluted from a RPC column as single-stranded amphipathic α-helices. In addition, the hydrophobic face of an amphipathic α-helix (made up primarily of hydrophobic residues in the ***a*** and ***d*** positions of the repeating heptad) will exhibit preferential binding (a "preferred binding domain" or pbd) to the reversed-phase packing compared to its hydrophilic face (Zhou *et al.,* 1990; Sereda *et al.,* 1995; Mant *et al.,* 1998). Hence, as has been demonstrated previously, any substitutions in the hydrophobic face of such an amphipathic peptide will have considerably more effect on the RPC elution behavior of that peptide compared to substitutions made in its hydrophilic face. To take this a step further, the relative retention times of amphipathic peptides with the 3-4, 4-3 hydrophobic repeat characteristic of the single α-helical strands of coiled coils may also reflect the subsequent relative stabilities of the resulting coiled coils.

The reversed-phase retention times of all peptide analogues are reported in Table 3-2 although, for the present, we are only concerned with the peptides exhibiting mutations in the hydrophobic core of the cassette region, i.e., I20A and I20T. Fig. 3-11 (top) shows the reversed-phase elution profile of a mixture of reduced peptides I20T and I20A together with the native and control peptides; these retention time data were then compared to the relative stabilities of their respective oxidized coiled coils (Fig. 3-11, bottom). From Fig. 3-11 (bottom; black dots), it can be seen that there was, indeed, a good correlation between observed reduced peptide retention time (an experimental measure of overall hydrophobicity of each helix making up the coiled coil) and coiled coil stability, reflecting previous observations by this laboratory with other *de novo* designed model coiled coils. Note from Fig. 3-11 (bottom; white dots) that, interestingly, the correlation between coiled coil stability and peptide hydrophobicity is also quite satisfactory when said hydrophobicity is expressed as the sum of hydrophobicity coefficients (developed from RPC-derived data) only of side chains in the dimer interface (see Fig. 3-5) of the cassette region, thus underlining the importance of the hydrophobicity of this central cassette region to the stability of the entire coiled coil.

III) Conclusions

In the long term, SCM should provide us with a new understanding of the role of long-range interactions in protein folding as well as offering insights into the

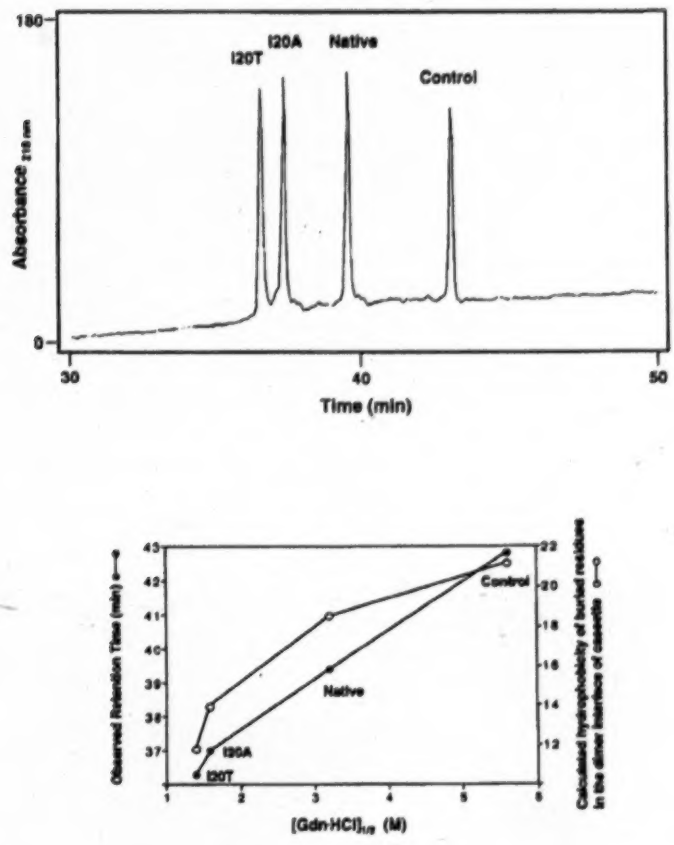


Fig. 3-11. Top: RPC of single stranded α-helical coiled coils. Bottom: Relationship between hydrophobicity of the reduced cassette holder containing control, native, I20T, and I20A sequences and their stability.

role of protein stability in inducing and controlling protein conformation. In the present study to determine the feasibility of the promising SCM approach, we have already demonstrated that the overall stability of the cassette holder is critical in nucleating the conformational change of the cassette; indeed, our data suggest that the nucleating sequences on either side of the cassette are critical in controlling the conformation of the cassette. Investigations into the role of secondary structure propensity of the central cassette region, as well as of residue hydrophobicity in the hydrophobic core of this cassette, so far suggest that a balance of the correct α-helical propensity and β-sheet propensity, as well as residue hydrophobicity at positions of potential interhelix or inter-β-strand hydrophobic interactions, is required for a sequence to adopt either conformation. Thus, immediate future development of the SCM model includes modifications to the secondary structure propensity and/or hydrophobicity of the nucleating helices in order to delineate further the nature and extent of long-range interactions that have such an important role in modulating protein structure and stability. In addition, the effect of cassettes containing other β-strands, β-turns, loops, regions of undefined structure, and helical segments on conformation and stability of our model protein will also be determined.

Chapter IV

Importance of secondary structural specificity determinants in protein folding: insertion of a native β-sheet sequence into an α-helical coiled-coil

Stanley C. Kwok, Colin T. Mant and Robert S. Hodges

A version of this chapter has been published in Protein Science (2002) 11, 1519-1531.

I) Introduction

Though the three-dimensional structure of a protein is determined by its amino acid sequence, the difficulty in understanding protein folding is unravelling the role each amino acid plays in determining the final fold and in stabilizing that fold. Any given residue in a protein may participate in a multitude of interactions, including those localized within a secondary structural element as well as longer-range contacts with other regions of the protein a considerable distance away in the amino acid sequence. A further complication lies in the inherent degeneracy in the coding of protein folding since a specific protein fold can be encoded by different amino acid sequences. For example, in the T4 lysozyme, the buried hydrophobic residues are largely interchangeable with other non-polar amino acids, causing little structural perturbation (Ohmura *et al.,* 2001; Xu *et al.,* 1998). Thus, an emerging

hypothesis is that not all residues are equally important for protein folding, i.e., some amino acid interactions are the determinants of a particular folding pathway, while others merely serve to stabilize the specified fold (Dill, 1999; Lattman & Rose, 1993). Another example of degeneracy in protein folding can be found when comparing thermophilic homologues of mesophilic enzymes, where multiple surfaced-exposed amino acid mutations enhance protein stability with little or no change in overall enzyme architecture (Lehmann *et al.* 2000). Therefore, we need to identify the "structurally-important residues", i.e., residues that dictate the protein fold. Distinguishing the determinants of protein folding is difficult because the folding process is context-dependent. For example, the local propensity of most short peptide sequences to adopt a defined structure is very weak, since they have little structure in aqueous isolation outside of their native protein scaffold.

The structural ambivalence of short peptide sequences is illustrated by the identification of chameleon sequences, i.e., a segment of amino acids that can be stabilized as alternate secondary structures depending on protein environment i.e., they are structurally ambivalent (Minor & Kim, 1996; Meizi, 1998). Minor and Kim (1996) have designed an 11-residue peptide which is unstructured in a benign aqueous environment but, when inserted into different regions of the B1 domain of protein G (G_{B1}), can be stabilized either as an α-helix or a β-sheet depending on the non-local tertiary contacts. In an alternate approach, using a *de novo* designed "Structural Cassette Mutagenesis" model protein, our laboratory inserted an 11-residue β-sheet secondary structural element (cassette) from human immunoglobulin

Fab into an α-helical coiled-coil host protein (or cassette holder) (Kwok *et al.* 1998a). This β-sheet cassette, when inserted with optimal hydrophobic alignment with the coiled-coil hydrophobic 3-4 repeat of the host protein, folded into a fully helical conformation, illustrating the stabilization of the native β-strand sequence in a non-native α-helical secondary structure. These results and a historical statistical survey of Protein Databank (Argos, 1987) that showed sequence-similar penta-peptides can exist in unrelated tertiary structures in different proteins highlight the importance of secondary structure adaptation to its protein context. Interestingly, the conformational switching phenomenon is well established in protein misfolding diseases. For instance, the normal prion protein (PrP^{C}) is a normal isoform with high α-helical content, but the disease-causing β-sheet-rich (PrP^{Sc}) isoform is associated with formation of insoluble protein aggregates, leading to senile plaques (review in Cohen, 1999; Prusiner, 1997). Conformational transformation is also observed with Alzheimer's β-amyloid peptide (Aβ), where it generally exhibits α-helical content in organic solvents, but exhibits predominantly β-sheet structure in water.

Cregut *et al.* (1999) studied short- and long-range interactions by engineering mutant G_{B1} proteins, where the region corresponding to the native α-helix was replaced by different stable β-hairpin secondary structural elements. The resulting mutant proteins had decreased stability compared to wild-type protein (up to -5 kcal/mol), but all the 13-residue β-hairpin insertions were converted to an α-helical conformation (i.e., they exhibited chameleon-like behaviours) by the tertiary

hydrophobic interactions that favour the stabilization of helical secondary structure, despite the observation that these sequences form β-hairpins in their native environment. These authors subsequently concluded that, although local interactions play an important role in protein stability, they do not confer the specificity of folding, which was largely determined by non-local tertiary contacts. From these observations, a natural question arises: are there sequences that are specific for one type of secondary structure and cannot fold into a different secondary structure (i.e., non-chameleon sequences), thus serving as determinants of protein folding? Kammerer *et al.* (1998) suggested that an important determinant of folding of many coiled-coils is a "trigger sequence" where the coiled-coils would not fold without this sequence. However, Lee *et al.* (2001) subsequently demonstrated that a specific sequence is not essential for coiled-coil folding but rather a coiled-coil will fold when its overall stability exceeds a critical threshold. This does not rule out the possibility that very stable sequences can nucleate or trigger folding. These results lend support to our hypothesis that polypeptide sequences which are highly specific for one type of secondary structure do exist, containing features or "secondary structural specificity determinants" (SSS determinants) that prevent the sequence from folding into alternate secondary structures. In order to test such a hypothesis, we believe it is necessary to use a well-defined model protein which can test the contribution to protein folding of both short- and long-range interactions while minimizing or eliminating any ambiguity in interpretation of results which may arise from using more complex native proteins. Thus, our favoured approach to solving

the protein folding problem is to insert a small, defined secondary structural element into a larger host protein with different secondary structural propensity, followed by identification of which residues profoundly affect the structure and stability of the host protein. As described previously (Kwok, *et al.,* 1998a), our minimalist structural cassette approach is based on a *de novo* designed α-helical coiled-coil motif, recognized as one of nature's favourite ways of creating a dimerization motif (Hodges, 1996; Micklatcher & Chmielewski, 1999). A major advantage of such a model is that there is only one type of secondary structure present in the host protein (i.e., the α-helix) and the effects of single residue substitutions on protein folding and stability are more straightforward to quantify than mutations in native globular proteins. Extensive work in our laboratory and others has led to a good understanding of interactions which stabilize coiled-coils, including: stability contributions of the hydrophobic core positions ***a*** (Wagschal *et al.,* 1999) and ***d*** (Tripet, *et al.,* 2000) in the hydrophobic *gabcdef* heptad repeat characteristic of this motif (Hodges, 1981; 1996); polypeptide chain length effects on coiled-coil stability (Su *et al.,* 1994; Fairman *et al.,* 1995); intrinsic amino acid side-chain propensities for α-helix or β-sheet structure (O'Neil & Degrado, 1990; Lyu *et al.,* 1990; Chakrabartty *et al.,* 1994; Zhou *et al.,* 1994a; Monera *et al.,* 1995); helix capping and termination signals (Aurora & Rose, 1998; Lu *et al.,* 1999); hydrogen bonding (Borg *et al.,* 2001); lactam-bridge stabilization (Houston *et al.,* 1996; Kwok *et al.,* 2001); and side-chain rotamer entropic effects (Yu *et al.,* 1999; Penel & Doig, 2001). In addition, the three-dimensional structures of many natural helical protein motifs

have been characterized by both crystallography and NMR experiments (reviewed by Kohn *et al.,* 1997b), thus providing a wealth of structural data for comparison with experimental observation. The insertion of an amino acid sequence ("cassette"), which exhibited β-strand structure in its native protein, into our host coiled-coil protein results in an α-helix-β-sheet-α-helix (host-guest-host) chimeric protein arrangement and may result in three possible folding scenarios: **1)** the α-helices in the host protein induce or nucleate β-sheet cassette to fold into an α-helical conformation; **2)** the sequence of the β-sheet cassette prevents itself from being induced by the host protein into α-helical structure, but the nucleating α-helices of the host protein still fold; and **3)** the sequence of the β-sheet not only does not fold into an α-helical conformation, but is able to prevent the host protein from folding into α-helical structure. The present study extends our investigation of the potential of our SCM model to study short-range and long-range interactions in the stabilization of secondary structural folds. Specifically, we describe its efficacy in identifying and characterizing novel secondary structure specificity determinants as well as quantifying their contributions to protein stability.

II) Results

A) Design of the α-helical coiled-coil host protein as a cassette holder

A detailed description of the *de novo* design and application of the Structural Cassette Mutagenesis (SCM) model has been reported previously in Kwok *et al*

(1998a). To summarize, we had designed a very stable α-helical coiled-coil host protein with a repeating heptad sequence (***gabcdef***) of E-V-E-A-L-K-K, where the hydrophobic core is occupied by Val and Leu residues in positions ***a*** and ***d***, respectively, forming a 3-4 (or 4-3) hydrophobic repeat (Fig. 4-1). The N-terminal α-helical segment of the host protein contains an interchain disulfide bridge, due to a Cys-Gly-Gly linker sequence at the N-terminal of each polypeptide chain, this linker facilitating coiled-coil folding and increasing stability. In addition, this flexible linker eliminates the concentration dependent monomer-dimer equilibrium in the formation of a two-stranded coiled-coil, thus making chemical denaturation studies independent of protein concentration (Zhou *et al.,* 1992a). Furthermore, the disulfide bridge also ensures that the polypeptide chains are in-register and parallel. From Fig. 4-1, the N-terminal α-helical segment of the polypeptide chain contains three hydrophobes in the hydrophobic core and the C-terminal α-helical segment contains four hydrophobes, thus maintaining the continuity of the 3-4 hydrophobic repeat characteristic of coiled-coils throughout the cassette holder. Besides hydrophobic stabilization by large hydrophobes (Val and Leu in ***a*** and ***d*** positions, respectively), inter- and intra-chain ionic interactions were introduced to stabilize the coiled-coil. Inter-chain salt-bridges were engineered by placing Lys and Glu residues at positions ***e*** and ***g*** of the α-helical segments resulting in ionic stabilization due to electrostatic attractions ($i \rightarrow i' + 5$ or g to e') between negatively charged Glu side-chains and positively charged Lys side-chains. In addition to such inter-chain electrostatics for

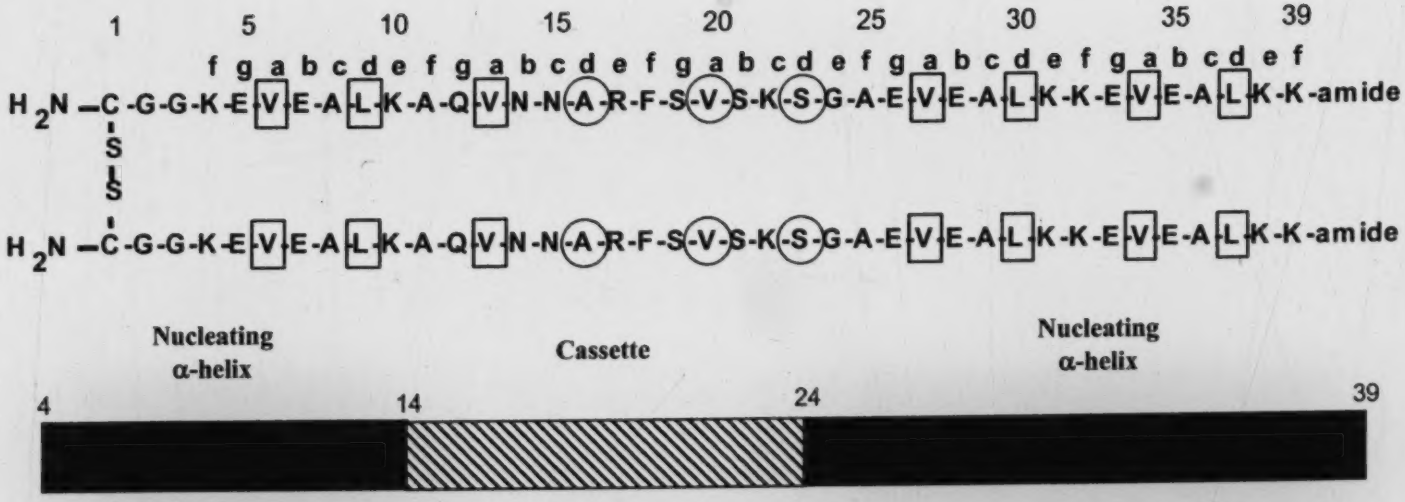

1 5 10 15 20 25 30 35 39
f g a b c d e f g a b c d e f g a b c d e f g a b c d e f g a b c d e f
H2N—C-G-G-K-E-V-E-A-L-K-A-Q-V-N-N-A-R-F-S-V-S-K-S-G-A-E-V-E-A-L-K-K-E-V-E-A-L-K-K-amide
S
S
H2N—C-G-G-K-E-V-E-A-L-K-A-Q-V-N-N-A-R-F-S-V-S-K-S-G-A-E-V-E-A-L-K-K-E-V-E-A-L-K-K-amide
Nucleating α-helix
Cassette
Nucleating α-helix
4
14
24
39

Fig. 4-1. Model of the 78-residue disulfide-bridged two-stranded coiled-coil host protein. This model peptide consists of two nucleating α-helices (residues 4-13, 25-39; open rectangles) at the N- and C-termini of each 39-residue polypeptide chain. The hydrophobic residues at positions *a* and *d* of the coiled-coil heptad repeat *abcdefg* (denoted above the polypeptide chain), forming the hydrophobic core of the two nucleating α-helices, are boxed. The 11-residue cassette (residues 14-24; hatched rectangle) is inserted into the centre of the cassette holder, and the residues in this immunoglobulin β-strand cassette (Fig. 4-2) that would be buried in the hydrophobic core if the entire cassette holder fold into a coiled-coil are circled.

stabilization of the coiled-coil, the potential to form intra-chain ionic interactions between Lys residues at either positions ***e*** or ***f*** to Glu at position ***b*** (*i* to *i* + 3 or *i* to *i* + 4) enhances the overall helix nucleation potential. Ala residues were chosen at positions ***c*** and ***f*** because Ala has the highest intrinsic helical propensity (Zhou *et al.*,1994), and a relatively small, non-bulky side chain that minimizes the introduction of extraneous context-dependent interactions with the cassette, for example Ala 11 and Ala 25. These helix stabilizing interactions were engineered to create a very stable coiled-coil host protein, where the α-helical segments have the potential to nucleate the induction of α-helical structure in the cassette region. Indeed, insertion of an 11-residue control cassette with a sequence identical to that of the cassette holder (E-A-L-K-K-E-V-E-A-L-K) into the centre of the host protein resulted in a very stable (GdnHCl denaturation midpoint of 5.6 M) and fully folded α-helical coiled-coil (the 11-residue cassette length was chosen as representative of the average length of α-helical and β-strand segments; Chou & Fasman, 1974a,b; 1978). This central region of the coiled-coil host protein was designed to be the region of cassette insertion, since perturbing the central region of a coiled-coil has been shown to have the greatest destabilizing effect on overall protein conformation (Zhou *et al.*, 1992b; Harbury *et al.*, 1993). In addition, insertion of a cassette into the central location permits optimal helix nucleation as the flanking coiled-coil regions have the potential to direct folding of the cassette sequence (in our case a β-sheet sequence) into helical structure from both ends.

The effectiveness of our model to induce α-helical structure in a peptide sequence originally exhibiting a β-sheet conformation in its native protein was previously described by Kwok *et al.* (1998a) for a β-strand sequence from immunoglobulin Fab, a protein consisting predominantly of β-sheet secondary structure. When this 11-residue β-strand sequence (7Fab:64-74) was inserted as a cassette into our model coiled-coil cassette holder, the entire chimeric protein folded into an α-helical coiled-coil conformation. Even the subsequent substitution of five Thr residues (Thr having a low α-helical propensity and the highest β-sheet propensity of the amino acids; Minor and Kim, 1994; Zhou *et al.*, 1994a) into the cassette, in addition to the three Thr residues already present, failed to overcome the strong α-helix nucleating effect of the cassette holder (Kwok *et al.*, 1998b). Thus, even this modified cassette sequence with significantly enhanced β-sheet potential was fully induced into α-helical structure by the nucleating α-helices. Although such a result might suggest that our model host protein has such strong α-helical potential that all β-sheet segments from native proteins may potentially be induced into α-helical structure, we hypothesized that specific sequences exist which cannot be induced in this way. Indeed, the identification of a peptide sequence which exhibited β-sheet structure in its native protein and which was shown either not to fold into an α-helix when inserted into our model cassette holder, or even to disrupt this very robust model which has been shown to be extremely effective in inducing α-helical structure, would be a sequence worthy of further study.

B) Selection of the parent cassette for insertion into the coiled-coil

As noted above, the first objective of the present study was to determine whether a native β-sheet sequence could be identified which can override the strong nucleating potential of the flanking α-helical coiled-coil regions of the cassette holder, thus illustrating an example of local propensity overriding the longer range tertiary stabilization of the coiled-coil 3-4 hydrophobic repeat. The second objective was to identify the sequence dependent features responsible for prevention of α-helix formation. Upon closer examination of the amino acid sequences in immunoglobulin Fab, an amphipathic β-strand (7Fab:53-63; Fig. 4-2) was identified which immediately precedes the "chameleon" β-sheet strand utilized as the cassette previously (Kwok *et al.,* 1998a). Interestingly, this sequence included an adjacent pair of Asn residues, Asn being a very common helix capping residue (Aurora and Rose, 1998). Thus, with this feature in mind, this β-strand sequence was chosen as our initial candidate cassette for the present study.

From Figure 4-2, in its native conformation in the Fab molecule, this β-strand buries two large hydrophobic residues in the hydrophobic core of the immunoglobulin fold (Phe 57 and Val 59) obviously to help stabilize the sequence in a β-sheet conformation. In a similar fashion, we optimized the alignment of the cassette to create the greatest possible burial of hydrophobic residues in the hydrophobic core of the coiled-coil if the sequence were to adopt an α-helical structure. Thus, the 3-4 hydrophobic repeat through the cassette would be as non-

Figure 4-2. Molscript drawing of residues 1-100 of immunoglobulin λ light chain fragment Fab New (PDB designation: 7FAB). The top panel shows the immunoglobulin fold with the β-strand (colored white) that was used as a cassette for insertion into the two-stranded α-helical coiled-coil cassette holder. The bottom panel is a side view of this strand showing the buried side-chains (colored white), Asn 54, Arg 56, Phe 57, Val 59 and Lys 61 and the solvent exposed side-chains (colored green), Asn 53, Ala 55, Ser 58, Ser 60 and Ser 62. The sequence of this 11-residue β-strand cassette (7FAB: 53-63) is shown in the bottom panel with the residues that would form a continuous 3-4 hydrophobic repeat with the cassette holder colored yellow. The orange arrows denote the solvent exposed residues. The host protein containing this sequence is referred to as the Parent peptide in Fig. 4-3 (residues 14-24).

polar as possible. Significantly, if this sequence was induced into an α-helix by the nucleating α-helical segments of the cassette holder, a small hydrophobe (Ala55) and a larger hydrophobe (Val 59) would occupy positions 16***d*** and 20***a***, respectively, thus maintaining the 3-4 hydrophobic repeat of the host protein, i.e., these two hydrophobes would be buried in the hydrophobic core of the potentially fully folded α-helical coiled-coil protein model. Thus, despite the presence of the polar Ser 62 from the Fab protein at hydrophobic position 23***d***, this cassette sequence alignment represents the highest potential for the strongly nucleating α-helices to induce this β-strand sequence (denoted Parent cassette; Fig. 4-3) into α-helical structure as well as maximize potential coiled-coil structure and stability.

C) Characterization of the host coiled-coil following insertion of parent cassette

To determine the effect of insertion of the Parent cassette into the cassette holder, we characterized the structure of our model protein using circular dichroism (CD). From Fig. 4-4, not only did the Parent cassette not fold into an α-helix under benign conditions when inserted into the cassette holder, but the nucleating helices of the host protein were also highly disrupted (only 23% overall helicity for the entire model protein; Table 4-1). Significantly, even in the presence of helix-inducing solvent (50% TFE), only 69% helicity was achieved, strongly suggesting the presence of features specific to the Parent cassette which are able not only to override the strong helix-inducing effects of the nucleating α-helices but the helix-

Peptide Number	Peptide Name	Amino Acid Sequence of Cassettes
		1 14 24 39
		c . . . . . b c d e f g a b c d e . . . . . f
1	Parent	H_2N- . . . . . N N A R F S V S K S G . . . . . -amide
2	S23L	H_2N- . . . . . N N A R F S V S K **L** G . . . . . -amide
3	N14A, N15A	H_2N- . . . . . **A A** A R F S V S K S G . . . . . -amide
4	N14A, N15A, S23A	H_2N- . . . . . **A A** A R F S V S K **A** G . . . . . -amide
5	N14A, N15A, S23L	H_2N- . . . . . **A A** A R F S V S K **L** G . . . . . -amide
6	S19A, S21A, S23L	H_2N- . . . . . N N A R F A V A K **L** G . . . . . -amide
7	N14A, S23L	H_2N- . . . . . **A** N A R F S V S K **L** G . . . . . -amide
8	N15A, S23L	H_2N- . . . . . N **A** A R F S V S K **L** G . . . . . -amide

Figure. 4-3. Amino acid sequences of the cassettes that were inserted into the centre of the two-stranded α-helical coiled-coil host protein. The Parent peptide contains the β-sheet cassette from the immunoglobulin fold (Fig. 4-2). Peptide nomenclature is based on the position of substitutions. For example, S23L denotes a leucine replacement of serine at position 23 of the cassette. These substitutions are boxed. The heptad repeat is denoted as *abcdefg*, where positions *a* and *d* are the positions in the hydrophobic core of the coiled-coil. The dotted lines denote the nucleating α-helices of the host protein (Fig. 4-1).

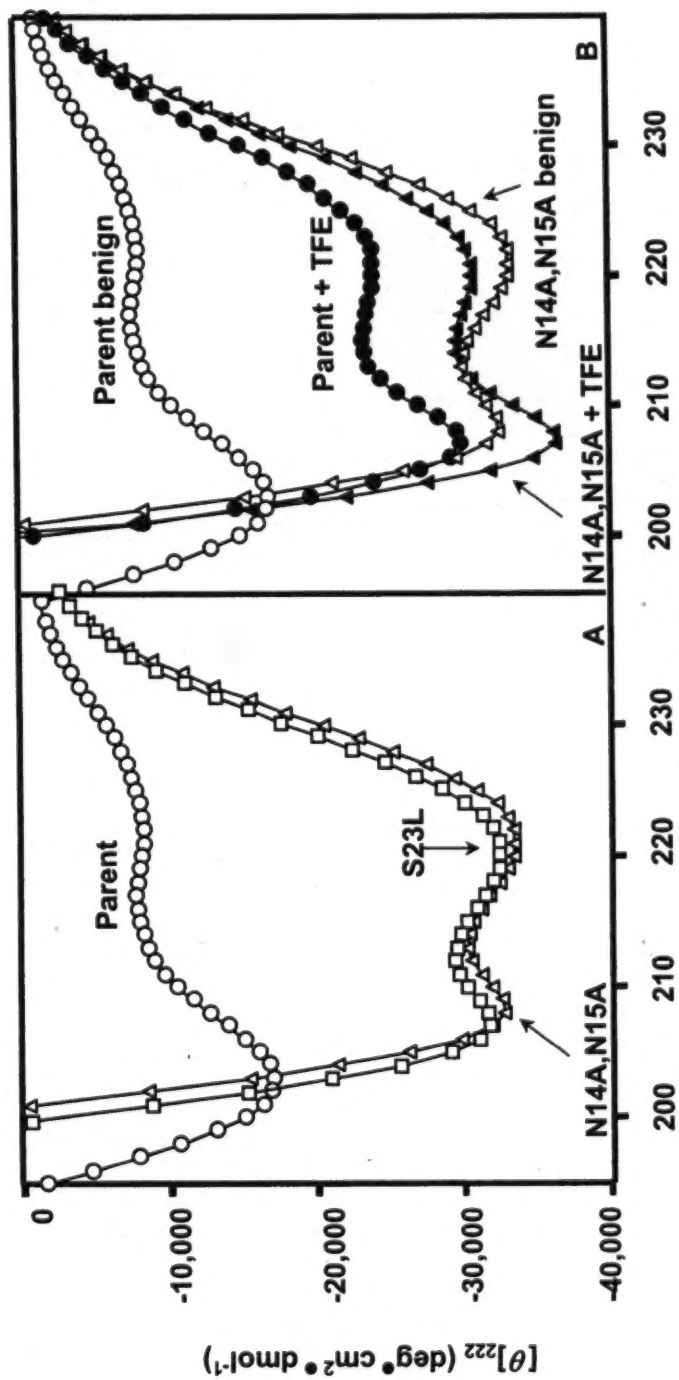


Figure. 4-4. Circular dichroism spectra of the cassette holder (Fig. IV-1) containing different cassette sequences at 25°C (Fig. 4-3). In Panel A, CD experiments were carried out in a 50mM PO_4 (K_2HPO_4/KH_2PO_4), 100mM KCl, pH 7.0 buffer with the following peptides: Parent (o), S23L (□), and N14A,N15A (△). In Panel B, CD scans of Parent and N14A,N15A peptides were carried out in a 50mM $KHPO_4$, 100mM KCl, pH 7.0 buffer in the absence (benign) or presence of 50% (v/v) trifluoroethanol (+TFE); Parent benign (o), Parent + TFE (•), N14A,N15A benign (△) and N14A,N15A + TFE (▲). Concentration of peptides ranges from 87 to 104mM.

Table 4-1. Circular Dichroism Results of the Oxidized Two-Stranded Peptides

Peptide Number	Peptide Name [a]	$[\theta]_{222}$ [b] Benign (25° C)	$[\theta]_{222}$ [b] 50% TFE	$[\theta]_{222/208}$ [c]	% Helix [d] Benign (25° C)	% Helix [d] 50% TFE (25° C)	$[Gdn^{\bullet}HCl]_{1/2}$ [e] (M)	m [f] (kcal•mol•M^{-1})	$\Delta\Delta G_u$ [g] (kcal•mol^{-1})
1	Parent	- 8,100	-24,100	0.64	23	69	- [h]	- [h]	- [h]
2	S23L	-32,800	-29,200	1.01	95	89	1.50	1.70	-1.8
3	N14A, N15A	-33,400	-30,600	1.02	96	88	1.00	1.67	-2.6
4	N14A, N15A, S23A	-33,100	-31,400	1.03	95	90	1.30	1.73	-2.2
5	N14A, N15A, S23L	-33,900	-34,100	1.03	97	99	2.55	1.73	0.0
6	S19A, S21A, S23L	-31,200	-32,000	1.01	90	92	1.80	1.68	-1.2
7	N14A, S23L	-32,600	-32,500	1.02	94	93	2.55	1.59	0.0
8	N15A, S23L	-33,200	-32,800	1.03	95	94	2.70	1.64	+0.3

[a] Peptide cassette sequences and nomenclature are shown in Fig. 4-3.

[b] $[\theta]_{222}$ is the mean residue molar ellipticity measured at 222 nm in a 50 mM $KHPO_4$, 100 mM KCl buffer, pH 7.0 in the absence (Benign) or presence of 50% trifluoroethanol (50% TFE) (v/v). Concentration of peptides ranges from 87 to 104uM. Uncertainty errors on the molar ellipticity reading is +/- 300.

[c] The helical ratio $[\theta]_{222/208}$ was calculated by dividing the observed molar ellipticity value at 222 nm ($[\theta]_{222}$) by the observed molar ellipticity value at 208 nm ($[\theta]_{208}$) in benign buffer.

[d] % Helix was calculated from $[\theta]_{222}$ based on an ellipticity value of -36,000 for 100% α-helical content derived from the equation $X_n^{\circ} = X_n^{\circ} (1-k/n)$, where X_n° is - 37,400, the wavelength dependent constant, k, is 2.5, and n is 36 residues (Chen *et al.*, 1974).

[e] $[Gdn\,HCl]_{1/2}$ is the denaturation midpoint of the two-state unfolding of an α-helical coiled-coil to a random coil. Guanidinium denaturation were carried out at 25 °C after overnight equilibration at room temperature. $[GdnHCl]_{1/2}$ is reproducible within +/- 0.05 M when the same stock solution of guanidinium hydrochloride was used. Values precised to 0.05 M were reported. Concentration of peptides ranges from 87 to 104uM.

[f] m is the slope term defined by the linear extrapolation equation $\Delta G_u = \Delta G_u(H_20) - m$ [denaturant], and $\Delta G_u(H_20)$ is the free energy of denaturation in the absence of denaturant, whereas ΔG_u is the free energy of unfolding at a given denaturant concentration.

[g] $\Delta\Delta G_u$, the change of the free energy of unfolding relative to peptide N14A,N15A,S23L , derived from the equation: $\Delta\Delta G_u = ([denaturant])_{1/2,x} - [denaturant]_{1/2,b})(m_u + m_b)/2$, (Serrano & Fersht, 1992; see Materials and Methods).

[h] Not determined because Parent peptide is mostly unfolded in benign conditions.

inducing properties of an excellent helix inducing solvent. Indeed, with the 11-residue Parent cassette representing about a third of overall polypeptide length, this helicity value indicates that two-thirds of the polypeptide sequence (the nucleating α-helices) are helical in the presence of 50% TFE, while the cassette retains random structure under these conditions.

Two regions of the β-strand sequence were deemed likely to be responsible for the prevention of helical structure in the cassette: **1)** the NN motif was suspect due to the low intrinsic helical propensity of Asn, in conjunction with the helix capping effects of Asn residues; and **2)** the unfavourable burial of a polar residue, Ser 23, in the hydrophobic core of the coiled-coil if the cassette was to fold into an α-helical structure. Thus, these two features were now studied further.

D) Identification of potential determinants that prevent α-helix formation in the cassette

Two analogs of the Parent sequence were now synthesized to delineate the potential contribution of the two regions noted above to the prevention of α-helix induction: **1)** S23L, where the polar Ser residue in position 23 of the model protein is replaced by a hydrophobic Leu residue, thus maintaining a continuous 3-4 hydrophobic repeat throughout the entire length of the protein sequence; and **2)** N14A, N15A, where there is a tandem replacement of two residues with low helical propensity (Asn) with two residues of high α-helical propensity (Ala). The

sequences of these analogs are shown in Fig. 4-3 (peptides 2 and 3). From Table 4-1, the S23L analog (peptide 2) forms a fully folded α-helical coiled-coil under benign conditions, i.e., the cassette holder has induced an α-helical conformation in the S23L cassette sequence, the entire protein now exhibiting a $[\theta]_{222/208}$ ratio >1.0 indicative of coiled-coil formation (Lau *et al.,* 1984; Yu *et al.,* 1996), 95% helicity and a molar ellipticity of –32,800° (Figure 4-4A; Table 4-1) which is not enhanced further by the addition of 50% TFE. Thus, the replacement of a single polar residue (Ser) with a hydrophobic residue (Leu) at position 23***d*** (Fig. 4-3) has had a profound impact on overall folding of the model protein, converting an essentially unstructured molecule (Fig. 4-4A; Table 4-1) into an α-helical coiled-coil with a chemical denaturation midpoint, $[GdnHCl]_{1/2}$, of 1.50 M (Figure 4-5B; Table 4-1).

In a similar manner to the S23L analog, N14A, N15A (peptide 3, Table 4-1) also exhibited a fully folded α-helical coiled-coil structure under benign conditions ($[\theta]_{222/208}$=1.02; Table 4-1), with 96% α-helicity and a molar elliplicity of –33,400° (Fig. 4-4A; Table 4-1) and a $[GdnHCl]_{1/2}$ value of 1.00 M (Fig. 4-5B, Table 1). Again, in a similar manner to S23L, no further induction of α-helicity is apparent with the addition of 50% TFE and a changed $[\theta]_{222/208}$ ratio <1.0 (0.84, Fig. 4-4B) for N14A, N15A is quite clear compared to that under benign conditions, indicating dissociation of the coiled-coil structure in TFE, a solvent known to induce α-helical secondary structure in potentially helical peptide sequences but to denature tertiary

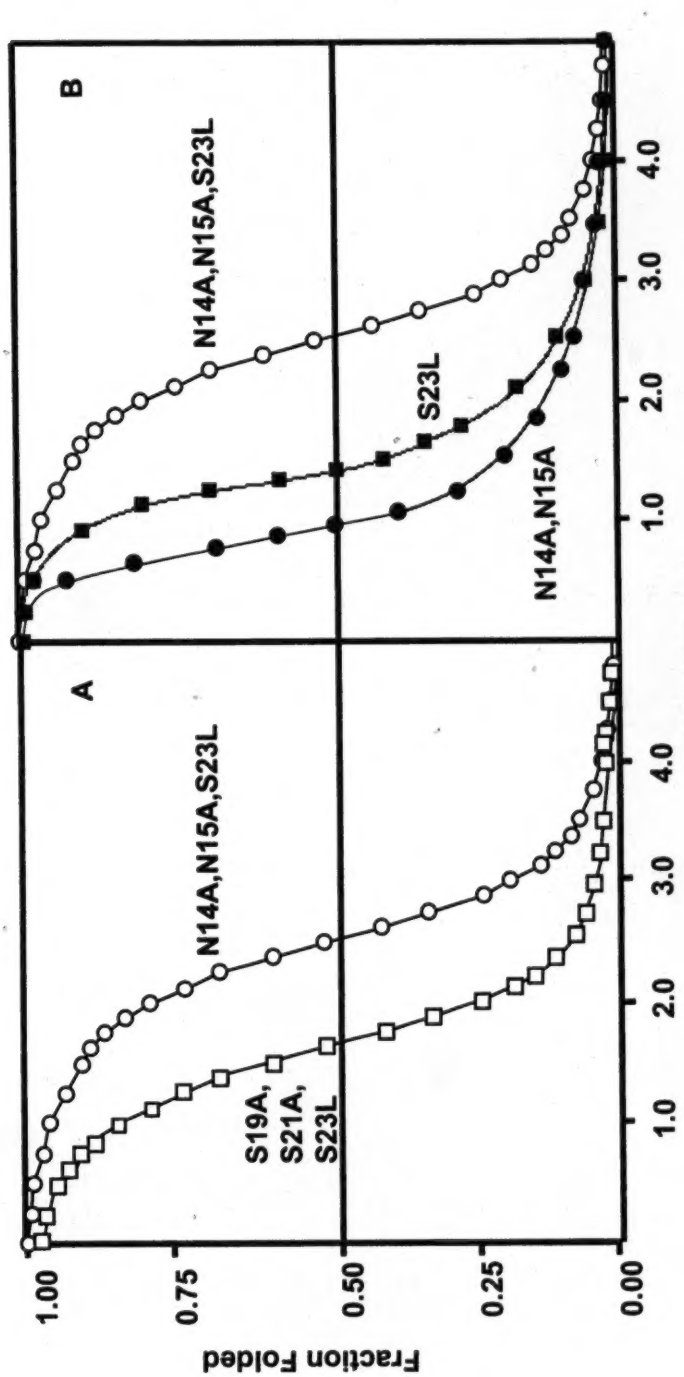


Fig. 4-5. GdnHCl denaturation profiles of selected peptide analogs. In Panels A and B, chemical denaturation experiments were carried out at 25°C in a 50mM PO_4 (K_2HPO_4/KH_2PO_4), 100mM KCl, pH 7.0 buffer with increasing concentrations of GdnHCl as denaturant. The denaturation dilution samples were vortexed and let to equilibrate overnight at room temperature. The fraction folded of each peptide was calculated as described in Materials and Methods. S23L (■), S19A,S21A,S23L (□), N14A,N15A (•), N14A,N15A, S23L (o). Concentration of peptides ranges from 87 to 104mM.

and quaternary structure (Cooper and Woody, 1990; Sönnichsen *et al.,* 1992). These results suggest that, when acting in concert, both the NN motif and Ser23 do indeed have a profound preventive effect on helix induction, i.e., they appear to be acting as secondary structure specificity determinants (SSS) by overcoming the strong helix inducing effect of the flanking nucleating α-helices. We now wanted to determine the relative contribution of the two regions to resistance to helix induction, as well as the requirement or otherwise of the NN motif, where Asn residues have their effect in tandem, as opposed to the Asn residues having their effect independently as single residues. Concomitant with these issues is what specific features of the two regions under scrutiny contribute to coiled-coil stability or instability when one or the other region has been removed. Although both the S23L and N14A, N15A analogs form fully folded α-helical coiled-coils under benign conditions (Fig. 4-4; Table 4-1), these molecules exhibit only moderate stability ($[GdnHCl]_{1/2}$ values of 1.50 M and 1.00 M, respectively; Table 4-1). New analogs would be designed to examine the quantitative effect of mutations in the NN motif and the hydrophobic core of the coiled-coil (residue 23***d***) on stabilization or destabilization of the coiled-coil.

E) Verification of the Asn-Asn motif as a helix destabilizing secondary structure specificity determinant

Two peptides analogs, N14A, S23L and N15A, S23L (Fig. 4-3) were designed to elucidate the role of the individual residues in the NN motif to the helix

disruptive properties of the Parent cassette. We used the S23L sequence to ensure folding of the cassette and increase stability of the coiled-coil to be able to measure substitution effects in the NN motif accurately. From Table 4-1, both analogs were fully folded α-helical coiled-coils, with similar $[GdnHCl]_{1/2}$ values of 2.55 M and 2.70 M for N14A, S23L and N15A, S23L, respectively, i.e., removal of either Asn residue significantly increased coiled-coil stability relative to S23L ($[GdnHCl]_{1/2}$ = 1.50 M; Table 4-1), where both Asn residues are present. These results support a synergistic contribution of both Asn residues to destabilizing effects on α-helical structure, as well as further emphasizing the role of NN motif as a helix destabilizing secondary structure specificity determinant (SSS).

F) Contribution of structural specificity determinant regions to coiled-coil Stability

Two analogs, N14A, N15A, S23L and N14A, N15A, S23A (Fig. 4-3), were designed to examine further the contribution to coiled-coil stability of residue hydrophobicity when substituted into the hydrophobic core at position 23***d***. Both analogs were folded α-helical coiled-coils under benign conditions, with $[GdnHCl]_{1/2}$ values of 2.55 M (N14A, N15A, S23L; Table 4-1, Fig. 4-3) and 1.30 M (N14A, N15A, S23A). Thus, the stability of the analogs increased with increasing hydrophobicity of the residue substituted into the hydrophobic interface at position 23***d***: from a $[GdnHCl]_{1/2}$ value of 1.00 M for the polar Ser at this position (N14A, N15A; Fig. 4-5, Table 4-1) to 1.30 M for the minimally hydrophobic Ala (N14A,

N15A, S23A; Table 4-1) to 2.55 M for the highly hydrophobic Leu (N14A, N15A, S23L). Aside from the expected coiled-coil stabilizing effect of increasing residue hydrophobicity at the non-polar 23***d*** position (Tripet *et al.*, 2000), the destabilizing effect of a polar residue (Ser in this case) is also quite clear. Interestingly, the stability of N14A, N15A, S23L appeared to correlate with the additive (and in concert) destabilizing effects of the unfavourable Ser23 burial in the coiled-coil core and the NN motif, with the S23L and N14A, N15A analogs exhibiting $[GdnHCl]_{1/2}$ values of just 1.50 M and 1.00 M, respectively (Table 4-1).

Finally, although it has been verified that the NN motif is indeed a secondary structural specificity determinant that resists induction of the Parent cassette into helical structure, the tandem nature of the NN dipeptide makes it more difficult to assess its destabilizing contribution to overall coiled-coil stability since the replacement of two Asn residues with two Ala residues (e.g., N14A, N15A, S23L) changes two factors simultaneously: **1)** the NN motif is removed, and **2)** there is an overall increase in helical propensity since Ala has a significantly higher α-helical propensity compared to Asn and it has been shown previously that amino acid α-helical propensity is an important factor governing protein stability (Monera *et al.*, 1995). Thus, in order to differentiate the destabilizing contribution attributed to amino acid propensity changes (as a result of the two Asn to two Ala replacements) from the intrinsic helix-destabilizing nature of the NN motif, analog S19A, S21A, S23L (Fig. 4-3) was prepared. Since Ser and Asn have the same intrinsic α-helical propensity (Zhou *et al.*, 1994), this analog should contain the same overall α-helical

propensity as that of N14A, N15A, S23L (note that, similar to the two Asn residues, neither of the Ser residues substituted at positions 19 and 21 of S19A, S21A, S23L are situated in the hydrophobic core of the coiled-coil). From Table 4-1, the S19A, S21A, S23L analog forms a fully folded α-helical coiled-coil under benign conditions, with a $[GdnHCl]_{1/2}$ value of 1.80 M (Table 4-1; Fig. 4-5). As is clear from Fig. 4-5, this stability value for S19A, S21A, S23L is considerably less than that of N14A, N15A, S23L, ($[GdnHCl]_{1/2}$ = 2.55 M; Table IV-1) where the highly destabilizing NN motif has been removed although overall helical propensity is identical, i.e., the difference in stability between the two analogs (N14A, N15A, S23L – S19A, S21A, S23L = 2.55 M – 1.80 M = 0.75 M change in the $[GdnHCl]_{1/2}$ values) may be attributed to the intrinsic destabilizing contribution of the NN motif separate from any helix propensity concerns.

III) Discussion

The folding of secondary structure generally correlates with overall protein hydrophobic stabilization and local folding preferences. However, our Parent peptide (Fig. 4-1) did not fold because sequence characteristics of the inserted β-sheet cassette prevented helix induction by the flanking α-helices in the host protein and the strong secondary structure specificity (SSS) determinants in the β-sheet cassette prevented the α-helical host from folding into a coiled-coil conformation. Even in a helix-promoting environment (50% TFE), only 69% helicity (i.e., two-thirds) was

achieved by the Parent peptide, underlying the strong helix disruptive elements of the remaining third of the sequence, i.e., the inserted β-strand cassette. It is also worth noting that the β-sheet cassette could not nucleate β-sheet structure in the flanking peptide sequences of the cassette holder since there is no β-sheet preference in these sequences, hence the observed general unfolding of the whole protein. By systematic removal of the helix disruptive motifs of the Parent cassette, we have achieved a β-sheet to α-helix transition of the cassette, allowing the host protein to fold into a coiled-coil. This systematic removal of sequence features with potential helix disruptive (or helix prevention) characteristics demonstrated how two adjacent Asn residues ("NN" motif) and an unfavourable burial of a polar Ser residue at the hydrophobic core of a putative coiled-coil acted in concert to prevent helix induction by the nucleating α-helices at either end of the Parent cassette.

Table 4-2 offers a summary of the effects of systematic mutation of the Parent cassette on subsequent coiled-coil formation and stability. Thus, specific pairs of cassette analogs are compared to highlight the contribution of specific sequence substitutions to coiled-coil stability, reported either as the change in free energy ($\Delta\Delta G_u$) per coiled-coil or $\Delta\Delta G_u$ per substitution or interaction (data adapted from $\Delta\Delta G_u$ values reported in Table 4-1). In each comparison of a pair of analogs, the $\Delta\Delta G_u$ value per coiled-coil (i.e., the relative change of free energy of unfolding in the coiled-coil due to a particular substitution(s); Table 4-2) was calculated by subtracting the $\Delta\Delta G_u$ value (i.e., the change in free energy of unfolding relative to

Table 4-2. The Change in Free Energy for Different Non-covalent Interactions

Analog Comparison		Description of contribution in each comparison [a]	Number of substitutions or interactions per coiled-coil [b]	$\Delta\Delta G_u$ per coiled-coil [c] (kcal•mol^{-1})	$\Delta\Delta G_u$ per substitution or interaction [d] (kcal•mol^{-1})
More Stable	Less Stable				
N14A,N15A,S23A	N14A,N15A	Contribution of Ala in place of Ser in the hydrophobic core	2	0.4	0.2
N14A,N15A,S23L	N14A,N15A	Contribution of Leu in place of Ser in the hydrophobic core	2	2.6	1.3
N14A,N15A,S23L	N14A,N15A,S23A	Contribution of Leu in place of Ala in the hydrophobic core	2	2.2	1.1
S19A,S21A,S23L	S23L	Contribution of Ala in place of Ser (increase α-helical propensity)	4	0.5	0.1
N14A,N15A,S23L	S23L	Destabilizing effect of the NN motif (includes N ->A propensity effect)	2	1.8	0.9
N14A,N15A,S23L	S19A,S21A,S23L	Net destabilizing effect of the NN motif	2	1.2	0.6
N14A,N15A,S23L	N14A,S23L	Contribution of single Asn(N15) -> Ala substitution	2	0.0	0.0
N14A,N15A,S23L	N15A,S23L	Contribution of single Asn(N14) -> Ala substitution	2	-0.3	-0.1
N14A,S23L	S23L	Destabilizing effect of the NN motif (includes N ->A propensity effect)	2	1.8	0.9
N15A,S23L	S23L	Destabilizing effect of the NN motif (includes N ->A propensity effect)	2	2.0	1.0

[a] Physical description of the effects of substitution(s) on coiled-coil stability.
[b] Number of interactions in a two-stranded α-helical coiled-coil; there are two identical polypeptide chains that are joined by a disulfide bridge in our cassette model.
[c] The relative change of the free energy of unfolding in the coiled-coil due to the substitution(s).
[d] The relative change of the free energy of unfolding per substitution or interaction. This is obtained by dividing the $\Delta\Delta G_u$/coiled-coil by the number of substitution(s) or interaction(s) in the coiled-coil, and rounded off to the nearest 0.1 kcal/mol.

N14A, N15A, S23L; Table 4-1) of the less stable analog from the more stable analog. For example, when gauging the contribution of Ala in place of Ser in the hydrophobic core (Table 4-2), $\Delta\Delta G_u$ for N14A, N15A, S23A minus $\Delta\Delta G_u$ for N14A, N15A = -2.2 – (-2.6) (Table 4-1; numbers modified to one decimal point) = 0.4 kcal.mol^{-1} per coiled-coil or 0.4/2 = 0.2 kcal.mol^{-1} per substitution (Table 4-2), i.e., the substitution of Ala for Ser has contributed 0.2 kcal.mol^{-1} to enhancement of coiled-coil stability for each of the two Ser → Ala substitutions.

Protein core hydrophobic interactions represent a major contributor to protein stability and this contribution is especially apparent during the nucleation of β-sheet to α-helix structural transition of our Parent cassette following removal of SSS determinants. Thus, the strong helix nucleation potential of our coiled-coil model is attributed to the host protein (cassette holder), possessing a 3-4 hydrophobic repeat, characteristic of α-helical coiled-coils, comprised of large hydrophobic Val and Leu residues. Subsequent insertion of a cassette with proper alignment of hydrophobic residues to allow a continuous 3-4 hydrophobic repeat throughout the whole protein sequence then allowed an accurate assessment of the effect of mutations in the inserted cassette sequence which stabilized or destabilized the hydrophobic core. Thus, the experimentally observed stabilizing effect of a hydrophobic Leu replacement for a polar Ser at position 23***d*** was 2.6 kcal.mol^{-1} or 1.3 kcal.mol^{-1} for each Leu → Ser substitution (Table 4-2). This trend of stability enhancement by removal of Ser from the hydrophobic core was also observed by Wagschal *et al*

(1999a,b) who reported an increase in stability of their coiled-coil model of 2.9 kcal.mol^{-1} when Ser was replaced by Leu at a ***d*** position; in addition, an Ala to Leu replacement increased coiled-coil stability by 2.1 kcal.mol^{-1}. Thus, our experimental values of 2.6 kcal.mol^{-1} and 2.2 kcal.mol^{-1} for Ser → Leu and Ala → Leu substitutions, respectively, correlated well with these earlier reported values. Significantly, in T4 lysozyme (i.e., in a globular protein context), the reported loss in stability associated with a Leu to Ala substitution was 2.1 kcal.mol^{-1}, with this value considered to be a measure of the difference in the free energy of transfer of Leu and Ala from the solvent to the protein core (Xu *et al.,* 1998). However, the contribution of large-to-small hydrophobic residue substitution is a more complex issue in a globular protein than in a coiled-coil, because the resulting cavity formation from the loss of van der Waal's interactions can result in dramatic loss of globular protein stability. In the coiled-coil setting, the replacement of Leu with Ala would represent a net loss of 78 Å^2 of hydrophobic surface area. By multiplying this hydrophobic area by two because there are two polypeptide chains per coiled-coil, and then by an associated hydrophobic factor of 25 cal/Å^2 (Karplus, 1997), we estimated that the decrease of stability is 2.0 kcal/mol, a value that is in excellent agreement with our observed value of 2.2 kcal/mol (Table 4-2).

Although the helix destabilizing property of the NN motif has not previously been reported, this helix disruptive nature correlates with known physical properties of Asn. For instance, Asn is a very common C-terminal helix capping residue (Aurora & Rose, 1998), has a strong tendency to form solvent exposed turns in

connecting transmembrane helices (Monne *et al.,* 1999) and has been commonly observed in crystal structure to participate in hydrogen bonding interactions with residues *i* to *i* + 2 ahead in the sequence (Wan & Milner-White, 1999a,b). However, the present study clearly demonstrated that the helix destabilizing effect of the NN motif (1.2 kcal.mol^{-1} per coiled-coil, including helix propensity effect; Table 4-2) is significantly greater than the contributions of the individual Asn residues. Indeed, when the individual Asn residues in the NN motif were substituted by Ala, the resulting contributions to coiled-coil stability were negligible (0.0 & –0.3 kcal.mol^{-1} per coiled-coil for Asn 15 and Asn 14, respectively; Table 4-2). Thus, the adjacent Asn residues of the NN motif are working synergistically to contribute significant helix destabilization.

The presence of the polar Ser residue in the hydrophobic core of the coiled-coil contributes more to coiled-coil destabilization than the NN motif (compare 2.6 kcal.mol^{-1} increase in coiled-coil stability for a Leu to Ser substitution with just 1.2 kcal.mol^{-1} for the net destabilizing effect of the NN motif; Table 4-2). However, the contributions of both the NN motif and Ser 23 are each of major significance since, as noted above, only when working in concert is the β-strand cassette able to override the helix nucleating influence of the α-helical host protein. Furthermore, the differences in free energy between the most stable peptide (N14A, N15A, S23L) and the two analogs containing single Asn to Ala substitutions (N14A, S23L and N15A, S23L) were, as noted previously, negligible (-0.3 and 0.0 kcal.mol^{-1}, respectively; Table 4-2). Thus, if either of these Asn residues was removed from the NN motif, its

Table 4-3. Numerical Analysis of Secondary Structural Specificity Determinants

Non-redundant occurrence of NN motif in the PDB [a]	504
a) found in α-helical structure [b]	8.1%
b) found in β-sheet structure [b]	5.9%
c) found in loops, turns or undefined conformations [b]	86.0%
Frequency of serine found in hydrophobic core of coiled-coils [c]	11%

[a] Performed on a non-redundant version of the Brookhaven Protein Database with 1021 protein sequences.
[b] Secondary structures were determined by SEQSEE and VADAR.
[c] Normalized occurrence relative to leucine (=100) (Tripet *et al.*, 2000).

destabilizing influence is essentially abolished. These observed destabilizing tendencies of the NN motif and Ser burial in the hydrophobic core of the coiled-coil may reflect a more global destabilizing approach found in natural proteins. Table 4-3 reports the results of Protein Data bank search for the occurrence of Ser burial in the hydrophobic core of naturally occurring coiled-coils and NN motifs in α-helices. From Table 4-3, Ser occurs with a frequency of only 11% in the hydrophobic core of natural coiled-coils; this compares with a 100% frequency for Leu. This observed infrequency of Ser in the hydrophobic core of native coiled-coils is indicative of its destabilizing effects on this secondary structure, clearly underlined by the present study. Interestingly the majority of NN motifs were found to occur in loops, turns or undefined structure (86%), in β-sheet segments (5.9%) and α-helices (8%) (Table 4-3), i.e., nature favours the NN motif to occur outside of protein secondary structure and its infrequency in helices may correlate with its helix-destabilizing nature.

As a final comment, it is interesting to note that, in the native Fab protein, the β-strand chosen as the cassette in the Parent peptide for the present study (Fab:53-63) immediately preceded the "chameleon" sequence (Fab:64-74) chosen as the cassette in our previous study introducing our cassette mutagenesis model (Kwok *et al.* 1998a). As reported previously, while exhibiting β-sheet structure in the native protein, this sequence was fully induced into α-helical structure when inserted into the cassette holder, in distinct contrast to the cassette sequence used in the Parent peptide of the present study. Thus, it is tempting to speculate that the strong

resistance to helix induction of the Fab:53-63 sequence may influence to some extent the preference for β-sheet structure of the chameleon sequence Fab:64-74 in the native protein. Certainly, such a possibility is worthy of further investigation.

Secondary structural specificity (SSS) determinants are important characteristics of polypeptide sequence influencing protein folding because they are very selective for one type of sequence and prevent the folding of alternative protein structure. These determinants are also important because they can exert long-range effects on neighbouring sequences, in contrast with the current view that short sequences are generally structurally ambivalent (chameleon-like) and play only a minor role in dictating protein conformation. We have shown that a short 11-residue β-sheet sequence can contribute significantly to helix destabilization in the host protein, and can also have a long-range influence on overall protein structure. Future work to characterize other secondary structural specificity determinants would allow us to understand the intricate interplay of short- and long-range interactions in determining protein folding.

Chapter V

Clustering of large hydrophobes in the hydrophobic core of two-stranded coiled-coils controls protein folding and stability

Stanley C. Kwok, and Robert S. Hodges

A version of this chapter has been published in Journal of Biological Chemistry (2003) 278, 35248-25254

I) Introduction

Understanding protein folding remains a challenging problem: how does information encoded in the amino acid sequence translate into the three-dimensional structure necessary for protein function? Although hydrophobic interactions are generally accepted as the predominant source of free energy change that maintains the folded state, this non-specific stabilization does not describe how the "hydrophobic collapse" guides the formation of specific secondary structure (α-helices and β-sheets) in the final tertiary and quaternary structure in the native protein. The concomitant model suggests that the hydrophobic collapse restricts the conformation of the polypeptide chain into a "molten globule" thus facilitating secondary structure folding in this limited conformational context (Dill, 1990). Examples of hydrophobic interactions participating in the early events of protein folding are observed via stopped flow fluorescence and nuclear magnetic resonance

(NMR) studies in apomyogloblin (Yao *et al.,* 2000) and cytochrome C (Colon *et al.,* 1996), illustrating the importance of the packing of non-polar residues in stabilizing helix-helix interactions. Recently, non-polar residues have also been observed to form non-native hydrophobic clustering in denatured proteins (Shortle & Ackerman, 2001; Klein-Seetharaman *et al.,* 2002), and the authors postulated that non-native hydrophobic interactions can stabilize the long-range order of the protein scaffold, i.e., via an intermediate not observed in the folded state, thus indirectly guiding the extended polypeptide chain towards the correct native fold. Such an observation suggests that the amino acid sequence encodes for structural characteristics other than that of the native fold; in other words, the hydrophobic patterning in the sequence encodes the pathway that ultimately leads to the native functional state. Considering that hydrophobic interactions mediate protein folding both in the folded and the unfolded state, how does a cluster of non-polar residues contribute to stability - is the free energy derived from the burial of hydrophobic residues simply a sum of the energy derived from the removal of non-polar surface area from aqueous medium, or does hydrophobic clustering enhance stability via other mechanisms, such as enhancement in enthalpy, geometric packing and van der Waals interactions.

The two-stranded α-helical coiled-coil is the simplest protein fold consisting of two amphipathic α-helices wound around one another forming a left-handed super coil stabilized by hydrophobic burial (Hodges *et al.,* 1990; Zhou *et al.,* 1992a,c). All coiled-coils share a characteristic heptad (seven-residue) repeat denoted as ***abcdefg*** in which non-polar residues occupy the ***a*** and ***d*** positions, forming an amphipathic

surface where non-polar interactions allowed assembly of two-, three- and higher oligomeric states (Lupas *et al.*, 1991). The quantitative contribution of 20 amino acids in positions ***a*** and ***d*** and their effects on protein stability and oligomerization state have been determined (Wagschal *et al.*, 1999a, 1999b; Tripet *et al.*, 2000). In addition, the secondary structure formation and hydrophobic collapse of coiled-coils are tightly coupled and very cooperative since single-stranded amphipathic α-helices are unstable in aqueous medium. This hydrophobic surface where amphipathic α-helices interact via hydrophobic interactions provide an ideal model to test the effects of hydrophobic clustering. We postulated that the hydrophobic clustering in the core of coiled-coils would have a significant influence on secondary structure formation and protein stability. Here we present the *de novo* design and characterization of two two-stranded α-helical coiled-coils that have the same inherent hydrophobicity, i.e., the identical hydrophobic core residues (6 Leu, 3 Ile and 7 Ala residues), but with different clustering of large and small hydrophobic core residues (Fig. 5-1). Their biophysical characteristics are compared by circular dichroism spectroscopy (CD), analytical ultracentrifugation, and differential scanning calorimetry (DSC). The results are discussed in the context of non-polar residue clustering enhancing protein stability.

Peptide Name	Amino Acid Sequences
	gabcdef
P3	Ac-E**A**EAL**K**A-E**I**EAL**K**A-K**A**EA**A**EG-K**A**EAL**E**G-K**I**EAL**E**G-K**A**EA**A**EG-K**A**EAL**E**G-E**I**EAL**K**A-GGCY-am
P2	Ac-E**A**EAL**K**A-E**A**EAL**K**A-K**I**EA**A**EG-K**A**EAL**E**G-K**I**EAL**E**G-K**A**EA**A**EG-K**A**EAL**E**G-E**I**EAL**K**A-GGCY-am

Peptide Name	Hydrophobic Clusters	Schematic Representation of Hydrophobic residues at *a* and *d positions*
		a d a d a d a d a d a d a d a d
P3	3 Clusters	
P2	2 Clusters	

Fig. 5-1. Peptide nomenclature, sequences and schematics of the coiled-coil analogs used in this study. Top panel shows the sequence of two peptides, P3 and P2 which contain 3 and 2 clusters, respectively. The heptad repeat is denoted as ***gabcdef*** which is shown above the first heptad of the coiled-coil. Bottom panel shows a schematic representation of the hydrophobic residues in positions ***a*** and ***d*** where large hydrophobes are shown by dark circles (Leu, Ile) and small hydrophobes are shown by open circles (Ala). Three continuous large hydrophobic residues in the core positions ***d, a, d*** constitute a hydrophobic cluster (rectangular box).

II) Results

A) Design of the α-helical coiled-coils with different hydrophobic clustering

The peptides used in this study were modeled on heptad sequences that had strong α-helical potential and a heptad repeat (***gabcdef)*** that facilitates coiled-coil formation. In the design of these hydrophobic clustered peptides, we took advantages of the features of the successful α-helical coiled-coil models in our laboratory (Hodges *et al.,* 1990; Zhou *et al.,* 1992a,c; Wagschal *et al.,* 1999a, 199b, Tripet *et al.,* 2000) for example, complementary packing in the protein core (Kellis *et al.,* 1989), balance of charged residues across the coiled-coil interface in heptad positions ***e*** and ***g*** (Kohn *et al.,* 1997a; McClain *et al.,* 2001), and a flexible disulfide-bridge linkage (O'Shea *et al.,* 1989). The coiled-coil sequences consisted of eight heptads (56-residues) based on two repeating heptad sequences: ExEAxKA and KxEAxEG where positions **x** are hydrophobic core positions occupied by Ala, Ile or Leu in positions ***a*** or ***d*** (Fig. 5-1.). We defined a hydrophobic cluster as a consecutive string of three large non-polar residues (Ile or Leu) in the core positions of the coiled-coil, and in our coiled-coils, non-polar residues Leu, Ile, Leu in the consecutive ***d***, ***a***, ***d*** heptad positions constituted a stabilizing hydrophobic cluster. Our approach was to design two proteins with identical inherent hydrophobicity, i.e., identical number and character of non-polar residues in equivalent coiled-coil core positions but with different arrangement, with P3 having three hydrophobic clusters and P2 with two (Fig. 5-1, rectangular boxes). The N-terminal hydrophobic cluster

of P2 was disrupted by shuffling Ile and Ala in positions ***a*** while maintaining the same inherent identical non-polar residue character in the core positions as P3 (Fig. 5-1, bottom). We created two analogs with identical inherent hydrophobicity, but with a different clustering pattern. The hypothesis is that the hydrophobic clusters are independent units that contribute to coiled-coil stability and folding when separated along the coiled-coil chain by consecutive strings of three alanine residues (Fig. 5-1, open circles). Thus, this pattern of large and small non-polar core residues helped specify the contribution of hydrophobic clustering from overall hydrophobicity. Inter-chain and intra-chain ionic interactions were engineered by placing Lys and Glu at positions ***b, e*** and ***g*** resulting in ionic stabilization due to inter-chain electrostatic attractions (*i* to *i'* + 5 or ***g*** to ***e'***) and intra-chain ionic attractions (*i* to *i* +3 or *i* to *i* + 4). To promote coiled-coil formation, a C-terminal disulfide bridge, Gly-Gly-Cys linker was introduced to facilitate the formation of a parallel and in-register coiled-coil, and a single Tyr-residue allows for protein concentration determination by UV spectroscopy. The disulfide bridge was distant from the N-terminal hydrophobic cluster under investigation (Fig. 5-1).

B) Secondary structure characterization by circular dichroism

Circular dichroism is a sensitive of probe of secondary structural features, and this technique was used to detect the difference in helical content between the two clustered peptides. Reduced P3 and P2 peptides were helical in benign

condition (approximately 50% helical), but in the helix-promoting environment of 50% TFE (Sonnichsen *et al.,* 1992), significant helical structure were induced (> 96%, Table 5-1.). Though the amino acid sequences of these peptide have a strong underlying helical propensity, potential stabilizing ionic interactions and amphipathicity, in the reduced state, there is insufficient hydrophobic stabilization at the concentration of ≈ 200 μM (Table 5-1) to form a fully folded coiled-coil. An interchain disulfide-bridge has been shown to enhance coiled-coil folding and stability by eliminating the concentration-dependent monomer-dimer equilibrium (Zhou *et al.,* 1992b, 1993). In contrast to the reduced peptide, both the disulfide-bridged two-stranded coiled-coils P3 and P2 exhibited more helical secondary structure, and the P3 coiled-coil with three hydrophobic clusters was fully folded at room temperature (Fig. 5-2). The P2 coiled-coil with two hydrophobic clusters was only 64 % folded at 20 °C in benign buffer though more helicity was induced at 5 °C (Figure IV-3, Panel A). P3 coiled-coil showed a $[\theta]_{222/208} > 1.0$, indicative of a fully folded coiled-coil whereas that of P2 coiled-coil is less than one, thus showing some population of P2 may be in the single-stranded unfolded state (Figure 5-2). Thus without the presence of three hydrophobic clusters, P2 did not fully fold. In the helix promoting solvent 50% TFE, both the disulfide-bridged peptides showed nearly 100% helical content (Table 5-1).

Table 5-1 Biophysical Characterization of of the Oxidized and Reduced Hydrophobic Cluster Peptides

Peptide Name [a]	$[\theta]_{222}$ (reduced) [b]			$[\theta]_{222}$ (oxidized) [c]			Sedimentation Equilibrium	
	Benign (20° C)	50% TFE	$[\theta]_{222/208}$ [d]	Benign (20° C)	50% TFE	$[\theta]_{222/208}$ [d]	Apparent M.W. [e]	Oligomer Ratio [f]
P3	- 18,800	-35,100	0.74	- 34,200	-34,900	1.02	14,800	1.2
P2	- 16,500	-34,600	0.69	- 22,800	-35,200	0.97	13,700	1.1

[a] Peptides are named by the number of stabilizing hydrophobic clusters shown in Figure 5-1.

[b] $[\Theta]_{222}$ (reduced) is the mean residue molar ellipticity measured at 222 nm in a reducing buffer 2 mM dithioerythritol (DTE), 50 mM PO_4, 100 mM KCl buffer, pH 7.0 in the absence (Benign) or presence of 50% trifluoroethanol (TFE, v/v) at 20 °C. Peptide concentrations were 206 uM and 220 uM (of monomer) for P3 and P2, respectively.

[c] $[\Theta]_{222}$ (oxidized) is the mean residue molar ellipticity measured at 222 nm in a non-reducing buffer of 50 mM PO_4, 100 mM KCl buffer, pH 7.0 in the absence (Benign) or presence of 50% trifluoroethanol (TFE, v/v) at 20 °C. Peptide concentrhations were 103 uM and 110uM (of dimer) for P3 and P2 respectively.

[d] The helical ratio $[\Theta]_{222/208}$ was calculated by dividing the observed molar ellipticity value at 222 nm by that at 208 nm.

[e] Apparent molecular weight were determined from sedimentation equilibrium analyses of disulfide-bridged peptides.

[f] Oligomeric states were calculated by dividing the apparent molecular weight by the mass of the disulfide-bridged two-stranded peptide.

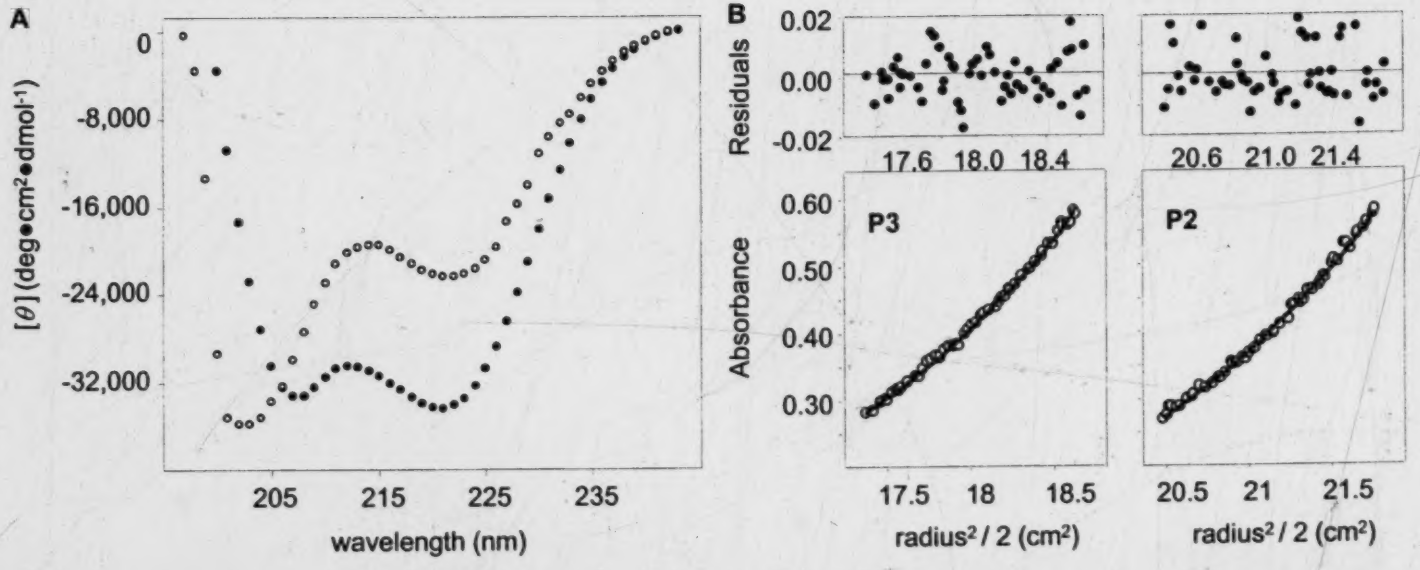


Fig. 5-2. Biophysical characterization of the two-stranded disulfide-bridged coiled-coil peptides: secondary structural content and oligomeric state. Panel A: Circular dichroism spectra of cluster peptides in benign condition (50 mM PO_4, 100 mM KCl, pH 7.0 at 20 °C). The concentration of P3 (•) and P2 (O) are 103 and 110 uM, respectively. Panel B: Sedimentation equilibrium of clusters peptides in benign condition at 20,000 rpm: the residuals are shown above the UV absorbance data which were best fitted to a single-species (two-stranded disulfide-bridged coiled-coil model). Initial peptide concentrations for P3 and P2 were 100 uM.

C) Sedimentation equilibrium analyses of oligomeric states of coiled-coils

The packing of hydrophobic core residues had been shown to affect the oligomerization states of coiled-coils (Wagschal *et al.,* 1999a, b). To determine that P2 and P3 had the same oligomerization state, sedimentation equilibrium analyses were carried out. Using different protein concentrations (from 50 – 500 μM) and rotor speeds (20,000 – 35,000 rpm), we found that the oligomerization behavior of coiled-coils P3 and P2 were best-fitted to that of a two-stranded coiled-coil; they sediment as single homogeneous species with molecular weights consistent with a two-stranded coiled-coil (Fig. 5-2 and Table 5-1). Overall, sedimentation equilibrium analyses showed that the interchange of hydrophobic residues of Ile and Ala that distinguish P2 and P3 did not change the overall oligomerization state, i.e. these coiled-coils are two-standed and do no assemble to high order oligomerization states. Therefore we attributed the difference in helicity to a difference in stability due to the presence of an additional stabilizing hydrophobic cluster in peptide P3.

D) Comparison of coiled-coil stabilities by thermal and urea denaturation

The P3 coiled-coil with three hydrophobic clusters exhibited a highly co-operative unfolding similar to that of native proteins (Fig. 5-3). In contrast, P2 coiled-coil was only partially folded with marginal stability. We determined the stability of these two analogs using thermal melting and urea denaturation, and found that P3 was significantly more stable than P2, with increase of thermal midpoint of

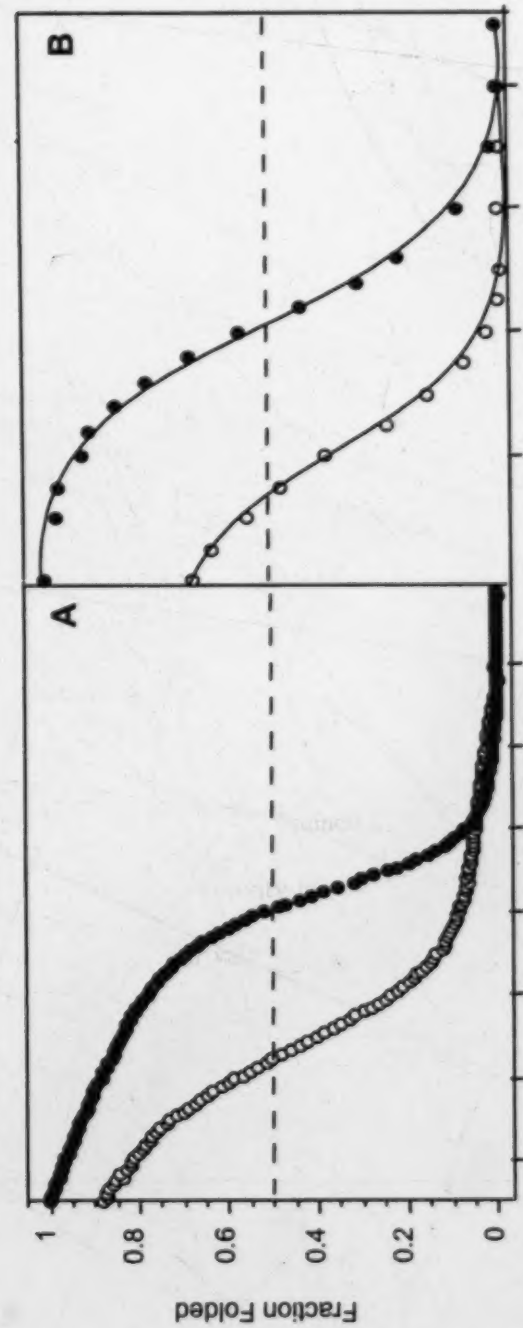


Fig. 5-3. Stability comparison of the P3 (3 cluster) and P2 (2 cluster) peptides. Panel A, temperature melting of the coiled-coils monitored by circular dichroism spectroscopy at 222 nm. Experimental condition was 50mM PO_4 (K_2HPO_4/KH_2PO_4), 100mM KCl, pH 7.0 buffer at a starting temperature of 5 °C. Panel B, urea denaturation experiments were carried out at 20°C in a 50mM PO_4 (K_2HPO_4/KH_2PO_4), 100 mM KCl, pH 7.0 buffer with increasing concentrations of urea as denaturant. Peptide P3 (•) and peptide P2 (○) had a respective concentrations of 103 and 110 uM of disulfide-bridged two-stranded coiled-coil. The ellipticity of peptide P3 was taken as fully-folded to determine fraction folded.

more than 18°C, and a corresponding increase in urea denaturation midpoint of approximately 1.5 M (Fig. 5-3, Table 5-2). To calculate the free energy difference between these two analogs, the linear extrapolation method was used to evaluate the chemical denaturation data (Table 5-2) because it is difficult to estimate thermodynamics parameter ΔH and ΔCp from non-ideal thermal melting curves (observed in P2). In addition, a more accurate protocol would be to estimate the free energy change from chemical denaturation, because the slope around urea denaturation transition is reproducible (Kellis *et al.*, 1989). The free energy estimated stabilizing contribution of N-terminal hydrophobic cluster in the P3 coiled-coil was 2.1 kcal•mol^{-1} per coiled-coil, and this difference in stability can explain why P2 coiled-coil did not fully fold. The slope term *m* associated with the transition for P3 was significantly higher than that of P2 (Table 5-2), and might describe the non-ideal unfolding behavior of the poorly formed secondary structure of P2. The *m* slope value generally correlates with the magnitude of changes in the non-polar accessible surface area between the folded and the unfolded state (Myers *et al.*, 1995).

E) Differential scanning calorimetry of coiled-coils

In addition to the CD experiments, DSC was also used to measure the contribution of hydrophobic clusters to coiled-coil stability. This technique is one of the most accurate methods for evaluation of protein unfolding transitions and analyses of enthalpy and heat capacity change of temperature-induced processes

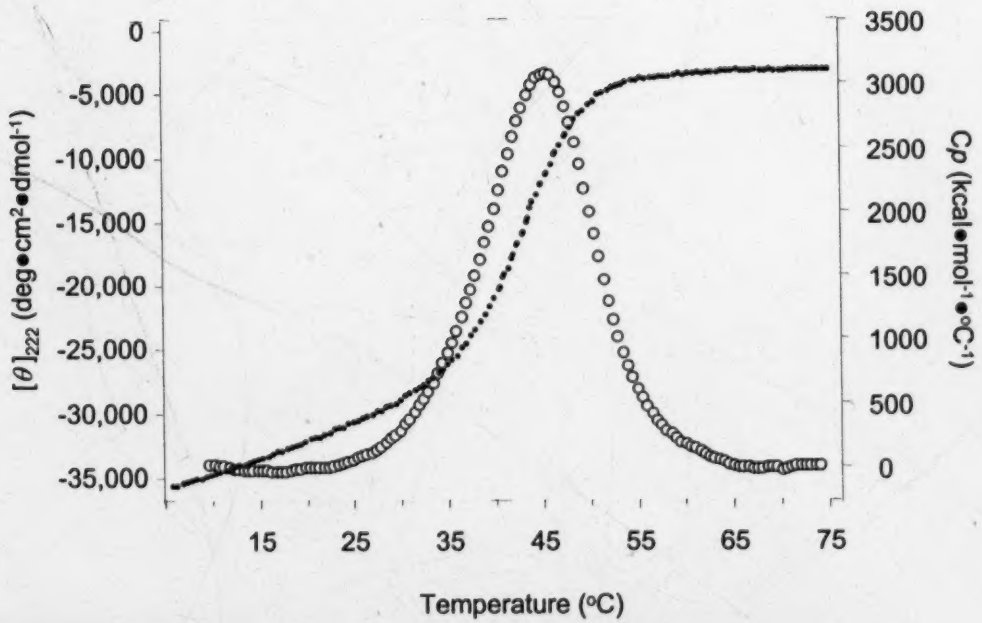


Fig 5-4. Stability profile of 3-clusters peptide: Thermal melt of P3 peptide as a function of temperature (•) overlayed on the differential scanning calorimetry profile (O). Experiment conditions was 50 mM PO_4, 100 mM KCl, pH 7.0 and peptide concentration was 103 uM of the disulfide-bridged two-stranded peptide.

(Privalov and Potekhin, 1986). The P3, the three-clustered peptide, showed a cooperative transition with a defined pre- and post-transitional baseline (Fig. 5-4), but the destabilized 2-clustered peptide, P2, did not exhibit any measurable transition peak (results not shown), thus illustrating the enthalpic and entropic disruption of the coiled-coil due to the loss of a hydrophobic cluster. Despite the fact that these two coiled-coils have the potential to bury nearly identical hydrophobic surface area, our results suggested that the two-state behavior of peptide unfolding/folding required a minimum of three hydrophobic clusters to form a stably folded protein (Fig. 5-4, $\Delta T_m \approx 42\ ^0C$). We then compared the thermodynamics parameter ΔH and T_m for P3 obtained from thermal melting and DSC experiments and found them to be in excellent agreement (Table 5-2). The estimated ΔC_p was 1.62 kcal•mol^{-1}•°C^{-1}, which equated to 13.5 cal•mol^{-1}•°C^{-1} per residue. This value was a relative measurement of the difference in exposed non-polar surface area between the folded and unfolded states (Pace, 1986; Xie & Freire, 1994) and agreed well with published values of those of small soluble proteins (Makhatadze & Privalov, 1995), typically in the range of 8-15 cal•mol^{-1}•°C^{-1} per residue.

III) Discussion

Hydrophobic interactions contribute significantly to protein stability because the burial of non-polar surface area is thermodynamically favorable in aqueous solution, and this study has shown that the hydrophobic stabilization is context

Table 5-2 Thermodynamics of Clustered Peptides and Stability Contribution of Hydrophobic Clustering

Peptide Name [a]	T_m (CD) [b] (°C)	T_m (DSC) [c] (°C)	ΔH (CD) [d] (kcal•mol^{-1})	ΔH (DSC) [e] (kcal•mol^{-1})	ΔC_p (DSC) [f] (kcal•mol^{-1}•C^{-1})
3 Clusters	41.6	42.3	40.0	42.9	1.62

Peptide Name [a]	T_m (melting) [b] (°C)	$[Urea]_{1/2}$ [g] (M)	*m*, slope [h] (kcal•mol^{-1}•M)	$\Delta G_{(H_2O)}$ [h] (kcal•mol^{-1})	$\Delta\Delta G_{water}$ [i] (kcal•mol^{-1})
3 Clusters	41.6	2.10	1.35	2.92	2.1
2 Clusters	23.1	0.65	0.67	0.86	-

[a] Peptides are named by the number of stabilizing hydrophobic clusters shown in Figure 5-1.

[b] T_m (CD) is the temperature at which there is a 50% decrease in molar ellipticity, $[\theta]_{222}$, compared to the fully-folded 56-residue coiled-coil peptide with 3 clusters as determined by circular dichroism spectroscopy.

[c] T_m (DSC) is the temperature midpoint associated with the transition peak during differential scanning calorimetry.

[d] ΔH (CD) is the change in enthalpy derived from the thermal denaturation experiment using CD spectroscopy.

[e] ΔH (DSC) is the change in enthalpy derived from the calorimetry experiment.

[f] ΔC_p is the change in heat capacity associated with the the denaturation transition observed in the differential scanning calorimetry experiment

[g] $[Urea]_{1/2}$ is the denaturation midpoint of the two-state unfolding of an α-helical coiled-coil to a random coil, i.e., the urea concentration (M) required to achieve a 50% decrease in molar ellipticity, $[\theta]_{222}$, with a fully-folded coiled-coil taken as –34,200 at 20 °C.

[h] *m* is the slope term defined by the equation $\Delta G_u = \Delta G_{(H_2O)} - m$ [denaturant], and $\Delta G_{(H_2O)}$ is the free energy of denaturation in the absence of denaturant whereas ΔG_u is the free energy of unfolding at a given denaturant concentration.

[i] $\Delta\Delta G_u$ is the free energy difference contributed by a stabilizing cluster: $\Delta G_{(H_2O)}$ (3 Clusters – 2 Clusters)

dependent. In the coiled-coil model, the additional cluster of three large non-polar residues in P3 enhances secondary structure folding and protein stability when compared to P2 in which this cluster is missing. Both the three-clustered P3 and the two-clustered P2 peptides have the same inherent hydrophobicity (6 Leu, 3 Ile, and 7 Ala residues in the hydrophobic core), yet their folding and stability differ dramatically. P3 with three hydrophobic clusters is a native protein-like two-stranded coiled-coil with a co-operative unfolding transition. In contrast, P2 with two hydrophobic clusters is only partially folded and significantly less stable compared to P3. Furthermore, P3 coiled-coil showed a single unfolding transition in DSC, whereas the unfolding transition was not measurable with P2 under identical experimental conditions. Thus, the disruption of the hydrophobic cluster in P2 drastically decreases enthalpy component of hydrophobic stabilization, which would affect the packing of the two interacting α-helices. The deconvolution of the entropic and enthalpic components of hydrobic-residue mutations, i.e., Leu to Ala mutation, is still controversial. For example, Dürr and Jelesarov (2000) found that the hydrophobic stabilization of a Leu to Ala substitution is large entropic at room temperature but becomes increasingly enthalpic at higher temperature.

Historically, hydrophobic cluster analysis structural prediction algorithm, based on the principles that hydrophobic amino acids cluster together in the native folded state, has been employed to good effect in identifying proteins with little sequence homology, but with similar overall protein topology (reviewed in Callebaut *et al.,* 1997). Although HCA does not predict the hydrophobic effect or protein

stability, our results clearly show that significant stabilization can be achieved when hydrophobes cluster in the coiled-coil core. Supporting the results from these prediction programs, experimental data on GCN4 coiled-coil folding suggested that the high-energy folding transition state is a hydrophobic collapsed form that contained little secondary structure (Sosnick *et al.,* 1996), and therefore, the hydrophobic effect can be an early folding determinant of protein folding, whereas formation of helical secondary structure occurs later.

Hydrophobic clustering may play a significant role in the structure and function of long native coiled-coil proteins. In a recent study on the assembly of tropomyosin, a 284-residue coiled-coil, Silva and co-workers observed that independent folding subdomains with different susceptibility to pressure were evident along its length (Suarez *et al.,* 2001). When subjected to increase pressure, tropomyosin melted in discrete cooperative blocks along the molecule, and the unfolded domains were likely unstable sites along the coiled-coil chain (Suarez *et al.,* 2001). We examined the hydrophobic clustering in the sequence of tropomyosin and indeed found 9 hydrophobic clusters occupied by 3, 4, 5 consecutive large hydrophobes (I, L, M, V, Y, F) in the core ***a*** and ***d*** positions, separated by destabilizing residues (A, S, T, Q, D, E and K) in the hydrophobic core (Fig. 5-5) (Hodges *et al.,* 1972). The authors of that study proposed that increasing osmotic pressure induced water molecules to infiltrate into the less stable "pressure-sensitive" domains in core of tropomyosin, but the more stable, "pressure-insensitive" domains maintain coiled-coil integrity. It is interesting, that our clusters

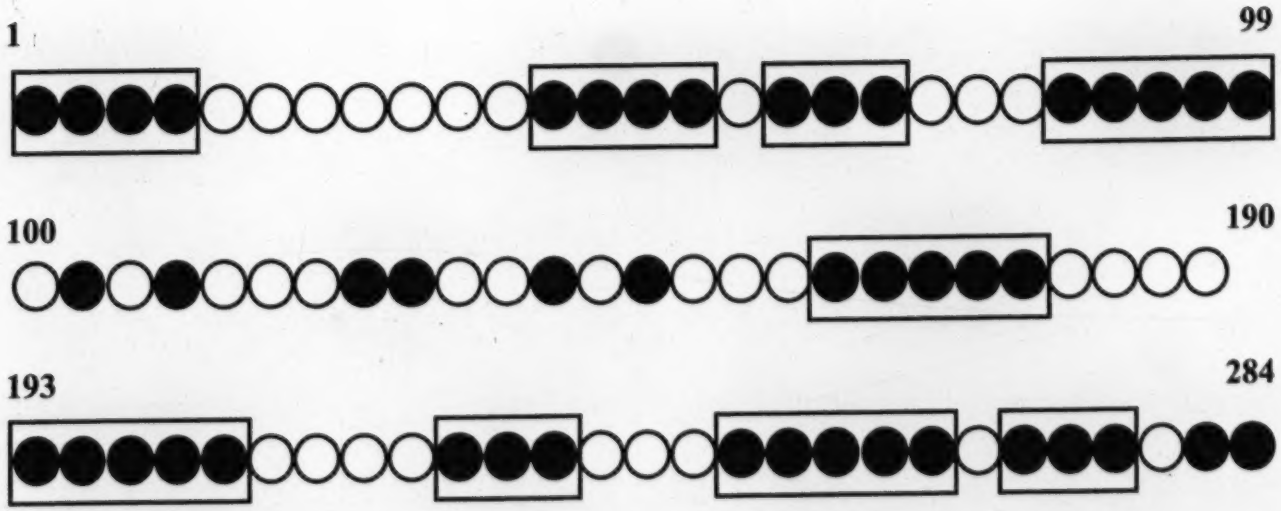


Fig. 5-5. Schematic representation of the hydrophobic core residues, in heptad positions ***a*** and ***d***, of the 284-residue coiled-coil protein tropomyosin. Large hydrophobic residues are shown by dark circles (Leu, Ile, Val, Met, Phe, Tyr) and small nonpolar, polar and charged residues are shown by open circles (Ala, Ser, Thr, Gln, Asp, Glu, and Lys)). Three or more continuous large hydrophobic residues in the core positions constitute a hydrophobic cluster (rectangular box).

of large hydrophobes (Ile, Leu) and small hydrophobes (Ala) could correspond to the "pressure-insensitive" and "pressure-sensitive" domains, respectively. Hydrophobic clustering may be an important mechanism for long native coiled-coil proteins to maintain chain integrity yet still accommodate the burial of polar and charged residues to control stability and facilitate different biological functions. Regardless of the length of the coiled-coil, hydrophobic clusters can serve as "knots" to keep the chain together while allowing flexible regions for function.

Hydrophobic clusters control protein stability, and are perhaps an evolutionary feature for native proteins to incorporate structurally-important stabilizing regions, as well as less stable and more flexible functional domains in a single protein fold. The destabilizing cluster regions of a coiled-coil may be involved in conformation change that allows for protein-protein interactions. For example, the less stable core regions of tropomyosin could be more easily disrupted for interactions with other proteins (members of the troponin complex and actin involved in muscle regulation) (Tripet *et al.*,1997). Another example of the importance of flexibility in coiled-coils is the so-called "spring-loaded" mechanism for viral fusion by the hemagglutinin protein of influenza (Carr & Kim, 1993), which allows viral entry into host cells. Conformational change in coiled-coil domains has been implicated, for example, in kinesin, kinesin-like, dynein motor proteins, intermediate filament proteins and tropomyosins (Tripet & Hodges, 2002; Chana *et al.*, 2002; Thomasin *et al.*, 2002; Araya *et al.*, 2002; Gee *et al.*, 1997; Gee & Vallee, 1998; Vallee & Gee, 1998) In addition to mediating protein interactions,

hydrophobic residues in the coiled-coil core have been known to control oligomerization state in *de novo* designed coiled-coils and maintain chain orientation in GCN4 by an Asn-Asn pairing (Gonzalez *et al.*, 1996). Thus destabilizing clusters could also be involved in these roles. Lastly, hydrophobic clusters are natural nucleation sites for protein folding, since they provide the necessary hydrophobic stabilization for early folding intermediates (Suarez *et al.*, 1996). Further investigations into the structural and functional roles of hydrophobic clustering will improve our understanding of the mechanism of coiled-coil folding and the folding of proteins in general.

Chapter VI

Stabilizing and destabilizing clusters in the hydrophobic core of long two-stranded α-helical coiled-coils

Stanley C. Kwok and Robert S. Hodges

A version of this chapter has been accepted for publication in the Journal of Biological Chemistry

I) Introduction

To decipher the problem of protein folding, a fundamental understanding of the stability contributions of non-covalent stabilizing and destabilizing interactions is needed. These interactions are encoded in the primary amino acid sequence and guide the initial hydrophobic collapse in an aqueous environment leading to a compact protein fold with well-defined secondary structures. The Levinthal's Principle infers that the protein folding process is not a random and tedious sampling of all possible conformation space (Levinthal, 1969), but rather specific pathways are initiated by native and non-native hydrophobic clusters of non-polar amino acids that maintain residual protein interactions even in experimental denaturing environments (Shortle & Ackerman, 2001; Klein-Seetharaman, 2002; Choy *et al.*, 2003). Furthermore, the local amphipathicity and hydrophobic clustering pattern of short

stretches of amino acid sequence often correlate with their secondary structure conformations, i.e. α-helices and β-sheets, thus many structural prediction programs incorporate hydrophobic cluster analysis into computational algorithms, leading to moderate success in helical content prediction (Xiong *et al.*, 1995; Callebaut *et al.*, 1997).

α-Helical assembly motifs are abundant in proteins and we have focused on what is probably the most wide-spread assembly motif, the dimeric α-helical coiled-coils as a model to study protein folding and stability (Kohn *et al.*, 1997a). The two-stranded coiled-coil is an ideal model for such studies because its rod-like nature makes the folding problem a one-dimensional problem, therefore removing much of the complexity found in globular proteins (Hodges, 1996). Because this protein motif is wide spread in nature (estimated that approximately 3% of all helical sequences in the Protein Databank are coiled-coils, Wolf *et al.*, 1997), the coiled-coil and the α-helix have served extensively as models for the study of interactions in proteins, for example, hydrophobic interactions (Zhou *et al.*, 1992a; Zhu *et al.*, 1993; Wagschal *et al.*, 1999a; Wagschal *et al.*, 1999b; Tripet *et al.*, 2000; Gonzalez *et al.*, 1996), ionic interactions (Kohn *et al.*, 1997a; Kohn *et al.*, 1998; Lyu *et al.*, 1992; Stellwagen *et al.*, 1992;Scholtz *et al.*,1993), chain orientation (Monera *et al.*, 1994; McClain *et al.*, 2001), chain length effects (Su *et al.*, 1994; Kwok & Hodges, 2004), secondary structure propensities (O'Neil & Degrado, 1990; Zhou *et al.*, 1993b, 1994a; Yang *et al.*, 1997), context dependent interactions such as helix capping motifs (Tripet & Hodges, 2002; Tidor, 1994; Lu *et al.*, 1999; Aurora & Rose, 1998),

and clustering of residues in the hydrophobic core (Kwok & Hodges, 2003; Lu & Hodges, 2004). Our basic approach is to use *de novo* designed coiled-coils to isolate specific interactions and test their significance to protein folding and stability.

Coiled-coils are characterized by a heptad repeat motif (***gabcdef***) in which residues in the hydrophobic core positions ***a*** and ***d*** are the primary determinant of folding initiation and stability (Wagschal *et al.,* 1999b; Tripet *et al.,* 2000). Therefore, the hydrophobic core of short coiled-coils (< 40 residues) is often highly enriched with bulky non-polar amino acid residues with few charged or polar residues. In contrast, sequence analysis of longer natural coiled-coil domains, such as the fibrous muscle coiled-coil tropomyosin (284-residues per polypeptide chain) showed that its core regions are more accommodating of small non-polar, polar or charged residues which make up approximately 40% of the core residues (Lu & Hodges, 2004).

Tropomyosin is an ideal protein for sequence analysis because it is a widely-distributed, sequence-conserved 100% α-helical coiled-coil with a continuous 3-4 or 4-3 hydrophobic repeat void of any helical stutters, stammers or breaks (Perry, 2001). When the tropomyosin core ***a*** and ***d*** residues is represented as a linear array, the continuity of large hydrophobes is often interrupted by small hydrophobic, polar or charged residues creating alternating groupings of what we describe as stabilizing and destabilizing clusters. Based on an experimental coiled-coil stability scale from Hodges and coworkers (Wagschal *et al.,* 1999b; Tripet *et al.,* 2000), who substituted all 20 natural amino acid residues into positions ***a*** and ***d*** in model coiled-coils, we

classified Leu, Ile, Val, Met, Phe and Tyr as stabilizing residues and the remaining 13 amino acids as destabilizing residues. Pro is not found in coiled-coils due to its disruption of α-helical structure.

Having defined the criteria of both destabilizing and stabilizing clusters based on free energy measurements, our laboratory described a detailed statistical survey of 34 full-length (284-residue) tropomyosins including different isoforms and 25 different species. The distribution of bulky, non-polar stabilizing residues versus destabilizing residues was not random, but rather, a cluster of three or more hydrophobic core residues are interspersed with clusters of alanine, hydrophilic polar or charged residues, creating alternating clusters of stabilizing and destabilizing residues within the protein (Lu & Hodges, 2004). In addition, despite the overall sequence homology among tropomyosin molecules, core residues within these clusters are even more conserved than those in non-cluster positions, suggesting that the cluster organization of core residues are perhaps conserved to provide critical structural and functional roles. In support of this concept that alternating regions of stability can exist along the coiled-coil, a recent study involving a pressure-induced denaturation experiment of tropomyosin showed that there are discrete domains of stability and instability along the tropomyosin molecule (Suarez *et al.*, 2001).

Despite the many truncation and sequence comparative studies on the tropomyosin molecule aimed to identify folding nucleation sites and critical regions for stability and function (Perry, 2001; Hitchcock-Degregori *et al.*,, 2002; Landis *et al.*, 1999; Paulucci *et al.*, 2002), there is no clear understanding of the relative

contribution of stabilizing and destabilizing clusters have on protein stability and what constitutes a destabilizing cluster. We speculate that the phenomenon of destabilizing and stabilizing clusters mediates important structural and functional roles in long native coiled-coils. In particular, the stabilizing clusters could maintain the inherent structure while allowing the less stable and more flexible regions to be involved in function.

Our goal in this manuscript was first to extend the sequence analysis of the distribution of stabilizing and destabilizing clusters to longer coiled-coils (>1000 residues) and then model the cluster interactions in de novo designed coiled-coils, which have been shown to be an excellent template (Micklatcher & Chmielewski, 1999; Regan & Jackson, 2003; Hill *et al.*, 2000) to evaluate folding and stability questions. We designed three series of coiled-coils to determine what defines a destabilizing cluster and varied the length of the destabilizing cluster from 3 to 7 residues in the hydrophobic core to determine the length effect of the destabilizing cluster on overall stability.

II) Results

Comparative analyses of tropomyosin sequences have been reviewed extensively in the literature (Perry, 2001), and the occurrence of Ala residues in destabilizing clusters in tropomyosin observed by our laboratory (Lu & Hodges, 2004) was especially relevant to the present research. The goal of this study was to

examine and compare the number and size of stabilizing and destabilizing clusters and intervening regions in long coiled-coils. The intervening regions do not involve clusters of stabilizing or destabilizing residues. We chose two representative coiled-coils for this analysis: tropomyosin of 284 residues per polypeptide chain and the domain of class II myosin heavy chain, an actin-based molecular motor that hydrolyzes ATP and converts chemical energy into mechanical force for movement (Sellers, 2000; Berg *et al.,* 2001; Weiss *et al.,* 1999), with >1086-residues per polypeptide chain in the coiled-coil domain. Because of its role as a motor protein, multiple copies of myosin genes are present in all organisms, and the conventional class II myosin is a highly conserved ubiquitous molecule with only short stretches of sequence diversity responsible for specific functions, thus making it useful for sequence analysis.

A) Stabilizing and destabilizing clusters in natural coiled-coils

In human skeletal myosin, 310 residues are involved in the hydrophobic core (***a*** and ***d*** positions), of which 130 are destabilizing residues and 180 are stabilizing, residues based on our classification of stabilizing residues as Leu, Ile, Val, Met, Phe and Tyr and destabilizing residues as Gly, Ala, Ser, Thr, Asn, Gln, Asp, Glu, Cys, Trp, His, Arg, Lys. Interestingly, these residues were grouped into stabilizing and destabilizing clusters which we defined as 3 or more consecutive residues in the ***a*** and ***d*** position of either stabilizing or destabilizing residues. Our analysis of myosin

resulted in the identification of 26 hydrophobic stabilizing clusters and 14 destabilizing clusters, representing 50.6% of the total number of core residues (Table 6-1). In comparison, 74.4% of core residues in tropomyosin were organized into 9 stabilizing clusters and 6 destabilizing ones (Table 6-1). Although the coiled-coil domain of myosin is nearly four times longer than in tropomyosin, the total numbers of stabilizing and destabilizing residues as a percentage of the total residues in the hydrophobic core were remarkably similar; 58.1% in myosin and 58.5% in tropomyosin were stabilizing residues, and correspondingly, the destabilizing residue composition was 41.9% in myosin and 41.5% in tropomyosin (Table 6-1). Furthermore, the percentage of destabilizing and stabilizing clusters compared to the total number of clusters was very similar. Myosin has 14 destabilizing clusters (35%) and 26 stabilizing clusters (65%). In comparison, tropomyosin has 6 destabilizing clusters (40%) and 9 stabilizing clusters (60%) (Table 6-1).

Table 6-1. Analysis of Clusters in Myosin and Tropomyosin

	Human skeletal Myosin	Human skeletal Tropomyosin
Total a and d positions	310	82
Total destabilizing residues	130 (41.9%)	34 (41.5%)
Total stabilizing residues	180 (58.1%)	48 (58.5%)
Total number of clusters and intervening regions [a]	62	22
Total number of clusters	40 (65%)	15 (68%)
Number of destabilizing clusters	14 (35%)	6 (40%)
Number of stabilizing clusters	26 (65%)	9 (60%)
Total number of intervening regions [a]	22 (35%)	7 (32%)
Total number of destabilizing residues in destabilizing clusters	50 [c]	23
Number of residues [b]	Ala 22 (44%) Thr 5 (10%)	Ala 17 (74%) Ser 3 (13%)

Total number of stabilizing residues in stabilizing clusters	107 [c]		38	
Number of residues [b]	Leu	50 (47%)	Leu	21 (55%)
	Ile	24 (22%)	Ile	6 (16%)
	Val	16 (15%)	Val	5 (13%)
	Met	11 (10%)	Tyr	4 (10%)
Total number of destabilizing residues in intervening regions	79 [c]		11	
Number of residues [b]	Lys	21 (27%)	Ala	6 (55%)
	Ala	10 (13%)	Gln	2 (18%)
	Glu	10 (13%)		
	Ser	8 (10%)		
	Thr	8 (10%)		
Total number of stabilizing residues in intervening regions	71 [c]		10	
Number of residues [b]	Leu	47 (66%)	Leu	4 (40%)
	Val	11 (15%)	Met	3 (30%)
			Ile	2 (20%)
			Tyr	1 (10%)

[a] Intervening regions within the hydrophobic core (***a*** and ***d*** positions) that do not contain destabilizing or stabilizing clusters
[b] Only the residues that occur at greater than 10% are listed.
[c] Two stabilizing residues are involved in frame shift regions; one destabilizing residue is involved in a frame shift region

Not surprisingly, the most popular residue in stabilizing clusters in both myosin and tropomyosin was Leu (~50%) (Table 6-1) which agreed with Leu being the most frequently found residue in the core of coiled-coils (Cohen & Parry, 1990; Lupas *et al.,* 1991). Interestingly, the destabilizing clusters in myosin and tropomyosin consisted mainly of Ala, where Ala was present in all six of the destabilizing clusters in tropomyosin and in twelve of fourteen of the destabilizing clusters in myosin (Table 6-2a and 2b). The high frequency of occurrence of Ala in the destabilizing cluster of myosin (44%, Table 6-1) and tropomyosin (74%, Table 6-1) is understandable considering the properties of Ala. Ala is the smallest non-polar amino acid side-chain, with the highest α-helical propensity of amino acid side chains (O'Neil & Degrado, 1990; Zhou *et al.,* 1993) and generally provides more stability to coiled-coils than any other residue classified as destabilizing (Wagschal *et al.,* 1999b; Tripet *et al.,* 2000). Thus, the smaller non-polar Ala allows for destabilization of the hydrophobic core but at the same time maintains the α-helical structure through its high α-helical propensity.

The size of hydrophobic clusters in tropomyosin varied from 3-5 residues, whereas in myosin the hydrophobic stabilizing clusters varied from 3-6 consecutive hydrophobes with the cluster length of 5 being most prevalent in tropomyosin and the cluster length of 4 being most prevalent in myosin (Table 6-2a and 6-2b). Though the destabilizing cluster length in tropomyosin and myosin varied from 3 to 7 consecutive destabilizing residues, a cluster length of 3 destabilizing residues was most common in both proteins (Table 6-2a and 6-2b).

Table 6-2a. Occurrence of Clusters in the Hydrophobic Core of Myosin

Human Skeletal Myosin Heavy Chain Coiled-Coil (residues 853-1938)[a]

Cluster Size	Sequence Positions	Residues in Stabilizing Clusters[b]	Sequence Positions	Residues in Destabilizing Clusters[b]
3	853, 856, 860	MMF	930, 933, 937	AEN
	877, 881, 884	LML	1185, 1189, 1192	AHT
	919, 923, 926	LIV	1460, 1464, 1467	WCT
	1091, 1094, 1098	MLI	1579, 1583, 1386	KKE
	1175, 1178, 1182	FML	1636, 1640, 1643	AAA
	1256, 1260, 1263	LLI	1699, 1703, 1706	TRA
	1671, 1675, 1678	LLV	1804, 1808, 1811	AAG
	1900, 1903, 1907	FIL	1833, 1837, 1840	ENA
			1910, 1914, 1917	AAA
4	1031, 1035, 1038, 1042	LVLL		
	1143, 1147, 1150, 1154	LLIL		
	1204, 1207, 1211, 1214	VLIL		
	1312, 1316, 1319, 1323	FILL	1129, 1133, 1136, 1140	ERAR
	1425, 1429, 1432, 1436	LVLV	1886, 1889, 1903, 1906	AASN
	1481, 1485, 1488, 1492	LLIY		
	1509, 1513, 1516, 1520	LILI		
	1622, 1626, 1629, 1633	MLML		
	1685, 1689, 1692, 1696	LILL		
	1790, 1794, 1797, 1801	LVLL		
5	954, 958, 961, 965, 968	LILLV	1157, 1161, 1164, 1168, 1171	ATQNR
	1228, 1232, 1235, 1239, 1242	MILMV	1600, 1604, 1607, 1611, 1614	STASD
	1562, 1565, 1569, 1572, 1576	IILVV	1752, 1755, 1759, 1762, 1766	AAAAA
	1734, 1738, 1741, 1745, 1740	LIIMI		
	1816, 1819, 1823, 1826, 1830	ILVLV		
6	979, 982, 986, 989, 993, 996	VLMLIL	-	
	1108, 1112, 1115, 1119, 1122, 1126	LILILI		
	1865, 1868, 1872, 1875, 1879, 1882	ILVLVY		

[a] Swiss-Prot accession number P12882.

[b] A stabilizing or destabilizing cluster is defined as 3 or more consecutive residues in the *a* and *d* positions of either stabilizing residues (I, L, V, M, F and Y) or destabilizing residues (G, A, C, S, T, N, Q, D, E, H, K, R, W). Stabilizing and destabilizing residues are based on the results of Wagschal et al., 1999 and Tripet et al., 2000.

Table 2b. Occurrence of Clusters in the Hydrophobic Core of Tropomyosin

Human Skeletal Tropomyosin Alpha Chain Coiled-Coil (residues 1-284)[a]				
Cluster Size	Sequence Positions	Residues in Stabilizing Clusters[b]	Sequence Positions	Residues in Destabilizing Clusters[b]
3	57, 60, 64 267, 270, 274	LYL YIL	74, 78, 81 116, 120, 123 151, 155, 158 235, 238, 242	AAA AAS AAA AAA
4	1, 4, 8, 11 39, 43, 46, 50 221, 225, 228, 232	MIML LLLL YILL	179, 183, 186, 190	AASC
5	85, 88, 92, 95, 99 162, 165, 169, 172, 176 193, 197, 200, 204, 207 246, 249, 253, 256, 260	VLIVL YVLIL LLVLL VLILL	-	
7	-		15, 18, 22, 25, 29, 32, 36	KAAAKAS

[a] Swiss-Prot accession number P09493

[b] A stabilizing or destabilizing cluster is defined as 3 or more consecutive residues in the *a* and *d* positions of either stabilizing residues (I, L, V, M, F and Y) or destabilizing residues (G, A, C, S, T, N, Q, D, E, H, K, R, W). Stabilizing and destabilizing residues are based on the results of Wagschal et al., 1999 and Tripet et al., 2000.

B) Intervening regions (outside clusters) in natural coiled-coils

We have defined intervening regions within the hydrophobic core (***a*** and ***d*** positions) as regions that do not contain destabilizing or stabilizing clusters. There are 22 intervening regions in the myosin hydrophobic core containing 71 stabilizing residues and 79 destabilizing residues. In tropomyosin, the number of intervening regions is 7 containing 10 stabilizing residues and 11 destabilizing residues. The ratio of stabilizing to destabilizing residues in intervening regions of both proteins, even through myosin is approximately 4 times longer, is the same and close to 1:1. Intervening regions in myosin varied in length from a single residue to as long as 27 hydrophobic core residues. However, 20 of the intervening regions varied from 2 to 13 hydrophobic core residues in length with 1 and 27 occurring only once. Intervening regions in tropomyosin vary in length from 1 to 7 hydrophobic core residues. Fig. 6-1 shows a map of the stabilizing and destabilizing clusters and intervening regions throughout the sequences of both proteins. The percentage of intervening regions to clusters was similar in both proteins: intervening regions in myosin 35% and tropomyosin 32%, with clusters making up 65% in myosin and 68% in tropomyosin (Table 6-1). What was different between the two proteins was that myosin has 48% of the hydrophobic core residues found in intervening regions whereas tropomyosin has only 26%. This difference is due to myosin having larger intervening regions than tropomyosin (Fig. 6-1). The stabilizing and destabilizing clusters and intervening regions were well distributed throughout the sequences of

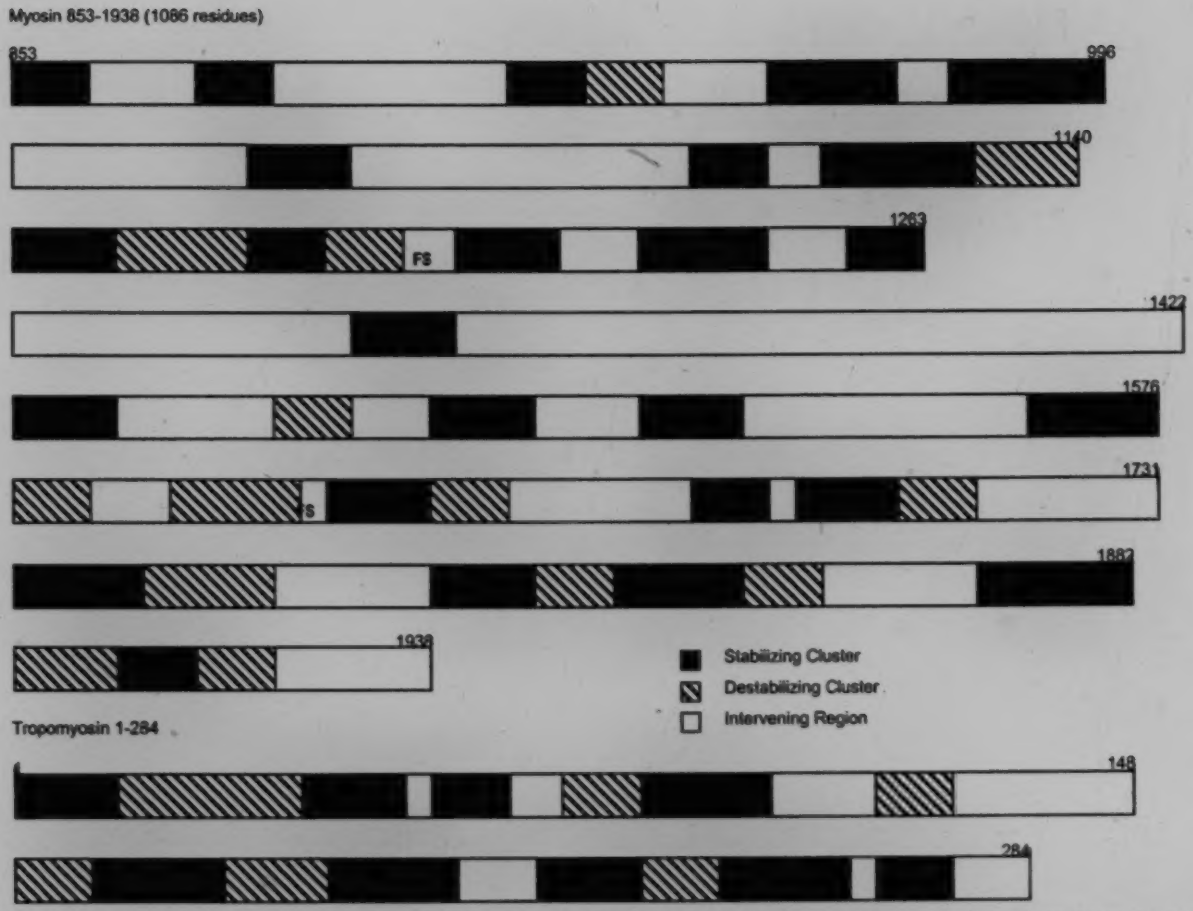


Fig. 6-1. Stabilizing and destabilizing clusters and intervening regions in two-stranded α-helical coiled-coils of human myosin (P12882: 853-1938, 1086 res.) and human tropomyosin (P09493: 1-284). The clusters and intervening regions are based on the composition of the hydrophobic core positions ***a*** and ***d*** of the heptad repeat $(\boldsymbol{abcdefg})_n$. When there is 3 or more consecutive hydrophobic core residues as stabilizing residues (Leu, Ile, Val, Met, Tyr or Phe), these regions are denoted as stabilizing clusters (black); or where there is 3 or more consecutive hydrophobic core residues as destabilizing residues (Gly, Ala, Asn, Gln, Ser, Thr, Asp, Glu, His, Arg, Lys, Cys, Trp) these regions are denoted as destabilizing clusters (hatched). Intervening regions (white) contain a mixture of stabilizing and destabilizing residues which are not in clusters. FS denotes a frame shift in the 3-4 or 4-3 hydrophobic repeat.

the two coiled-coils (Fig. 6-1). In tropomyosin, the dominating destabilizing residue found in destabilizing clusters and intervening regions was Ala (Table 6-1). In contrast, in myosin the dominating destabilizing residue found in the destabilizing clusters was Ala but in the intervening regions the dominating residue found in the hydrophobic core was Lys (Table 6-1). The dominating stabilizing residue found in the stabilizing clusters and intervening regions is Leu in both myosin and tropomyosin (Table 6-1).

C) Coiled-coil design

We designed a series of two-stranded parallel α-helical coiled-coils with different lengths (5, 6 and 7 heptads), all containing a N-terminal and C-terminal hydrophobic cluster (the C-terminal cluster has three consecutive core positions, whereas the N-terminal cluster has four consecutive core positions occupied by stabilizing large aliphatic hydrophobes, Ile and Leu at positions ***a*** and ***d***, respectively) (Fig. 6-2). Ile and Leu were shown to be major stabilizing residues in the hydrophobic core ***a*** and ***d*** positions relative to Ala. (Wagschal *et al.*, 1999b; Tripet *et al.*, 2000). Also, Ile at position ***a*** and Leu at position ***d*** were shown to favor the two-stranded coiled-coil over higher-order oligomers (Harbury *et al.*, 1993). In addition, these residues were also the most common ones in the stabilizing clusters of myosin and tropomyosin, making up 70% (Table 6-1). The α-heptad E-I-E-A-L-K-A was used at both ends of all coiled-coils. This provided in addition to the stable

Peptide Name	Heptad Code	Amino Acid Sequences
		1 . 7 14 21 28 35 42 49 53
		gabcdef
5A1	$\alpha_2\beta_2\alpha$	Ac-EIEALKA-EIEALKA-KAEALEG-KAEALEG-EIEALKA-GGCY-am
5A3	$\alpha_2\gamma\beta\alpha$	Ac-EIEALKA-EIEALKA-KAEAAEG-KAEALEG-EIEALKA-GGCY-am
6A1	$\alpha_2\beta_3\alpha$	Ac-EIEALKA-EIEALKA-KAEALEG-KAEALEG-KAEALEG-EIEALKA-GGCY-am
6A3	$\alpha_2\gamma\beta_2\alpha$	Ac-EIEALKA-EIEALKA-KAEAAEG-KAEALEG-KAEALEG-EIEALKA-GGCY-am
6A5	$\alpha_2\gamma_2\beta\alpha$	Ac-EIEALKA-EIEALKA-KAEAAEG-KAEAAEG-KAEALEG-EIEALKA-GGCY-am
7A1	$\alpha_2\beta_4\alpha$	Ac-EIEALKA-EIEALKA-KAEALEG-KAEALEG-KAEALEG-KAEALEG-EIEALKA-GGCY-am
7A3	$\alpha_2\gamma\beta_3\alpha$	Ac-EIEALKA-EIEALKA-KAEAAEG-KAEALEG-KAEALEG-KAEALEG-EIEALKA-GGCY-am
7A5	$\alpha_2\gamma_2\beta_2\alpha$	Ac-EIEALKA-EIEALKA-KAEAAEG-KAEAAEG-KAEALEG-KAEALEG-EIEALKA-GGCY-am
7A7	$\alpha_2\gamma_3\beta\alpha$	Ac-EIEALKA-EIEALKA-KAEAAEG-KAEAAEG-KAEAAEG-KAEALEG-EIEALKA-GGCY-am

Peptide Name	Number of Heptads	Number of Consecutive Ala(s)	Schematic Representation of Hydrophobic residues at ***a*** and ***d*** *positions*
			a d a d a d a d a d a d a d
5A1	5	1	
5A3	5	3	
6A1	6	1	
6A3	6	3	
6A5	6	5	
7A1	7	1	
7A3	7	3	
7A5	7	5	
7A7	7	7	

Fig. 6-2. Peptide nomenclature, sequences and schematics of the coiled-coil analogs used in this study. Top panel shows the three series of coiled-coil proteins, 5A1 and 5A3; 6A1, 6A3 and 6A5; and 7A1, 7A3, 7A5 and 7A7 which are denoted by the heptad length (5, 6 and 7) and by the length of the alanine cluster (A1 to A7). For example, 6A5 refers to a six heptad coiled-coil with a five-alanine cluster. Each analog is also named by the arrangement of its constituent heptad repeats described by the heptad code. Three different heptads were used (a: EIEALKA; b: KAEALEG; g: KAEAAEG). The alanine destabilizing cluster size is increased by replacing g-heptads for b-heptads that is a single Leu to Ala substitution which is denoted by the rectangle box (e.g. Leu in 5A1 is replaced by an Ala in 5A3). The heptad repeat is denoted as ***gabcdef*** which is shown above the first heptad of the coiled-coils. The residues at positions ***a*** and ***d*** in the hydrophobic core are bold. Bottom panel shows a schematic representation of the hydrophobic residues in positions ***a*** and ***d*** where large hydrophobes are shown by dark circles (●: Leu, Ile) and small hydrophobes are shown by open circles (O: Ala). The boxes indicate three or more consecutive Ala residues in the hydrophobic core. Na-acetyl, Ac; C-terminal amide, am.

hydrophobic core at positions ***a*** and ***d***, interchain and intrachain ionic interactions from Lys and Glu residues at positions ***b***, ***e***, and ***g***, resulting in ionic stabilization due to interchain electrostatic attractions (*i* to *i*′ + 5 or ***g*** to ***e***′) and intrachain ionic attractions (*i* to *i* + 3 or *i* to *i* + 4) (Kohn *et al.*, 1997, 1998). Ala at positions ***c*** and ***f*** provided excellent helical propensity, further increasing stability. As described previously (Kwok & Hodges, 2003), we introduced a C-terminal disulfide bridge, Gly-Gly-Cys-Tyr linker, to facilitate the formation of a parallel and in-register coiled-coil and the single Tyr residue per polypeptide chain allowed for protein concentration determination by UV spectroscopy. The disulfide-bridge also removed the monomer-dimer equilibrium observed in two-stranded coiled-coils making the comparisons of coiled-coil stability concentration independent (Zhou *et al.*, 1992b).

Based upon the above analyses, we chose to incorporate Ala as a destabilizing residue in the hydrophobic core of these coiled coils. The additional advantage of choosing Ala as a destabilizing residue is that it does not introduce any context dependent interactions such as hydrogen bonding or electrostatic effects that my occur with polar or charged side-chains. In Fig. 6-2 we have boxed the Leu to Ala substitutions that created three, five and seven residue Ala clusters in the hydrophobic core of these coiled-coils. For example, a single Leu to Ala substitution in peptide 7A1 created a three Ala cluster in 7A3; a single Leu to Ala substitution in 7A3 created a five Ala cluster in 7A5; and a single Leu to Ala substitution in 7A5

created a seven Ala cluster in 7A7. Otherwise the peptides are of identical length and sequence.

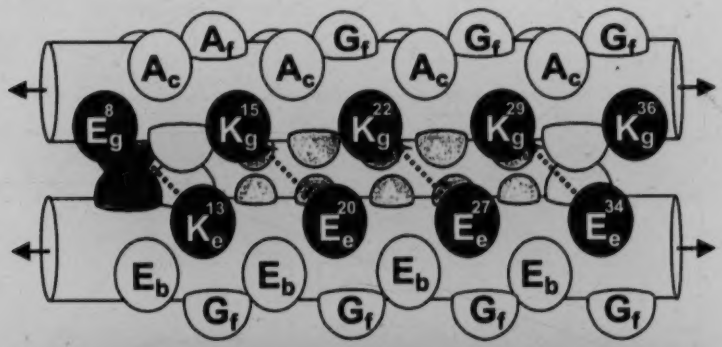
We designed three versions of two-stranded parallel α-helical coiled-coils with different heptad lengths (5, 6 and 7 heptads denoted as peptides 5, 6 and 7). The number of residues ranged from 39 to 53 residues. The coiled-coils differed in the number of consecutive Ala residues in the hydrophobic core from 1, 3, 5 and 7 (denoted A1, A3, A5 and A7). Thus, peptide 5A3 meant the peptide contained 5 heptads with 3 consecutive Ala residues in the hydrophobic core (Fig. 6-2). We chose different coiled-coil lengths to see if a Leu to Ala substitution was affected by the overall predicted stability of the coiled-coil. The starting coiled-coils, peptides 5A1, 6A1 and 7A1, were assured to have different stabilities since we deliberately added a β-heptad with lower overall stability than the α-heptad described above. The β-heptad, K-A-E-A-L-E-G, differed from the α-heptad in that the hydrophobic core consisted of Ala in position ***a*** instead of Ile. In addition, Gly was chosen to replace Ala at position ***f*** which has a much lower helical propensity than Ala and by switching the Lys at ***g*** and the Glu at ***e*** we maintained the same interchain electrostatic attractions i to $i' + 5$ but introduced i to $i + 4$ and i to $i + 3$ intrachain repulsions with the Glu residue at position ***b*** which further destabilized the heptad. Thus, peptides 5A1, 6A1, and 7A1 consisted of two, three and four β-heptads respectively. The helical propensity effect was estimated to be ~0.5 kcal/mol for each Ala to Gly substitution and 0.2 kcal/mol for the added repulsions (Zhou *et al.*, 1993) in each chain per heptad compared to the Ile to Ala substitution at position ***a***

of 1.95 kcal/mol for each substitution when made in the center of a hydrophobic cluster or 3.9 kcal/mol in a homodimeric coiled-coil (Wagschal *et al.*, 1999b; Tripet *et al.*, 2000). This suggests that the β-heptad could be as much as 5.3 kcal/mol less stabilizing than the α-heptad in a homodimeric coiled-coil (3.9+ 2*(0.5)+ 2*(0.2)= 5.3 kcal/mol). Therefore, as the chain length increased from 5 to 7 heptads, the overall coiled-coil stability was expected to decrease. The γ-heptad K-A-E-A-A-E-G was identical to the β-heptad with the exception that Ala replaced the Leu in the in β-heptad at position ***d***. This substitution was expected to decrease stability by as much as 1.90 kcal/mol for each substitution when made in the center of a hydrophobic cluster or 3.8 kcal/mol in a homodimeric coiled-coil (Tripet *et al.*, 2000). Thus, the γ-heptad could be as much as 3.8 kcal/mol less stabilizing than the β-heptad in a homodimeric coiled-coil. The single Leu to Ala substitution in each chain of the homodimeric coiled-coils used in this study would be expected to be significantly large and equal in all the coiled-coils unless there is a context dependency of the location of the Leu residue. For example, one could ask the question, does the location of a leucine residue next to a destabilizing cluster have a dramatic effect on its contribution to overall stability of the coiled-coil? To minimize context-dependent effects other than the desired effects under investigation (destabilizing clusters), the Leu to Ala substitutions were made in the center of the coiled-coils (Fig. 6-2), since these substitution sites should be less susceptible to end-fraying effects (Zhou *et al.*, 1992a).

A schematic representation of the coiled-coil 6A5 showed the spatial orientation of the destabilizing cluster and the conceptualization of small alanine residues relative to the larger hydrophobic residues Leu or Ile in the hydrophobic core (Top Panel, Fig. 6-3). Inter-chain attractive ionic-interactions mediated by charged residues at ***e*** and ***g*** provided additional stabilization across the coiled-coil interface. A space-filling model of the side-chain interactions of Leu-Leu and Ile-Ile pairing in the coiled-coil core (6IL) was modeled into an idealized coiled-coil domain of cortexillin (Bottom Panel Left, Fig. 6-3). On the right, coiled-coil analog 6A5 with a destabilizing cluster of five consecutive Ala was modeled into the same coiled-coil backbone showing the less than ideal van der Waals contact and spatial gap (Bottom Panel, Fig. 6-3). This leaves the hydrophobic core in the Ala cluster much more accessible to H_2O and competition with hydrogen bonds that stabilize the individual α-helices and thus an expected destabilization relative to Leu and Ile residues in the hydrophobic core.

D) Structural and biophysical characterization of coiled-coils

For the coiled-coils to be useful models that permit the investigation of a destabilizing cluster effect, they have to be dimeric, fully α-helical and exhibit reasonable stabilities when oxidized. The circular dichroism (CD) spectra of a representative oxidized coiled-coil, 5A1, showed excellent helicity, > 97%, in benign buffer characteristic of physiological pH and ionic strength, and even in the helix-promoting solvent 50% TFE (99% helix), no significant helix induction was



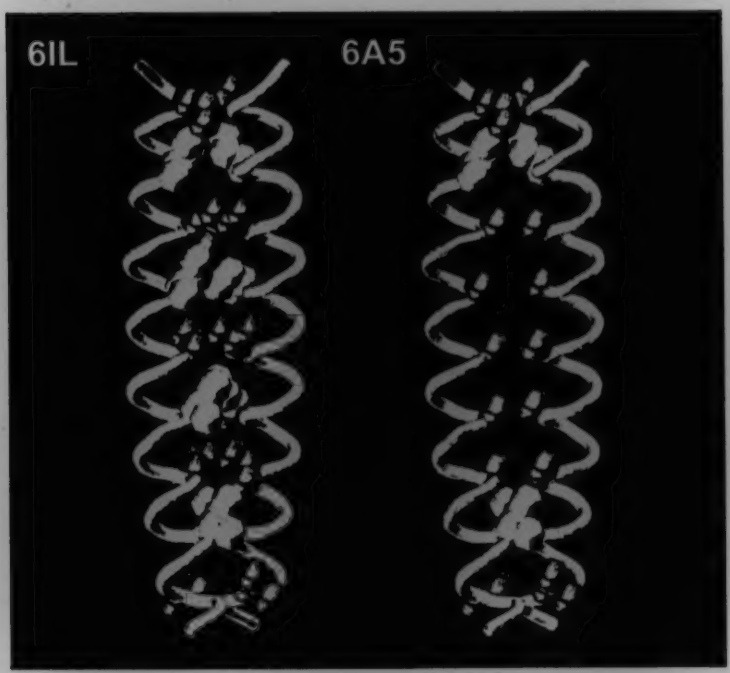


Fig. 6-3. Models of coiled-coils with either a stabilizing cluster (Ile, Leu) or a destabilizing cluster (Ala) in the hydrophobic core. (Top Panel). A schematic model of a five alanine-residue cluster located at the center of peptide 6A5 (Residues 8-36). The five small consecutive alanine residues (brown) on the non-polar surface of an amphipathic helix pack "knobs-into-holes" onto the alanines (brown) on the other amphipathic helix. The alanine residues are smaller compared to the other large hydrophobic residues (Ile, green and Leu, yellow). Inter-chain electrostatic interactions (*i* to *i'* +5) between Lys (red) at position *g* and Glu (blue) at position *e'* are shown by the dashed double-headed arrow. (Bottom Panel). A side-view of space-filling models depicting the different side chain packing interactions in a stabilizing clusters of Ile and Leu residues in the hydrophobic core (6IL) and a destabilizing cluster of five Ala residues (6A5) in the hydrophobic core, at positions *a* and *d*, of a model two-stranded parallel coiled-coil (the cortexillin dimerization domain, 1D7M was used to build the model). Side chains of Ile (green), Leu (yellow) and Ala (brown)

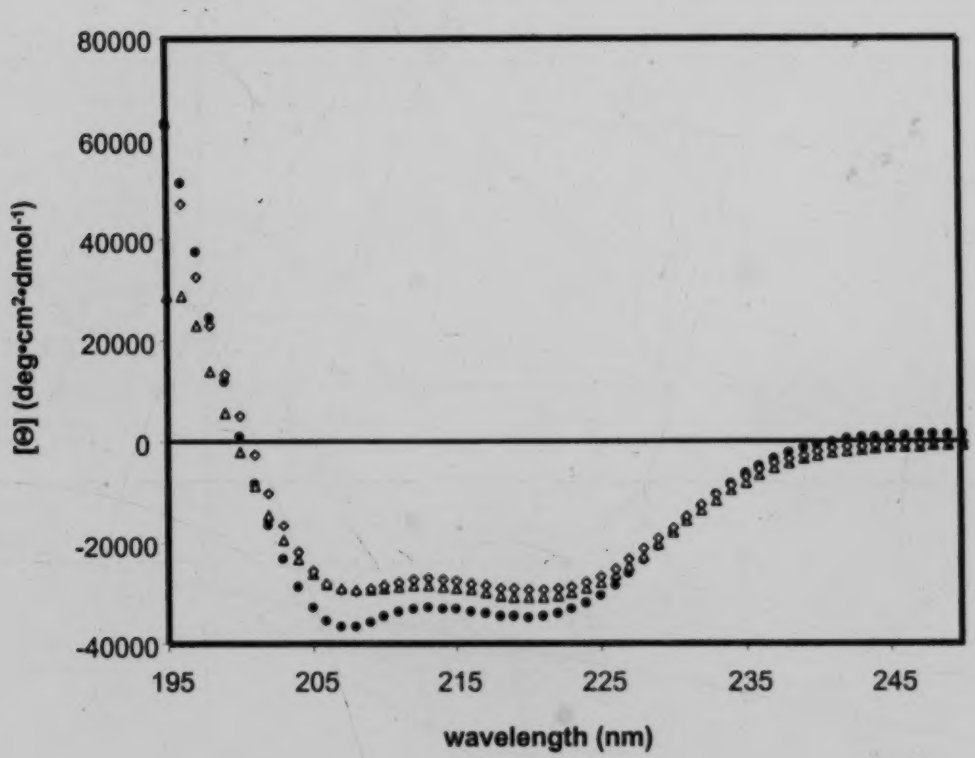


Fig. 6-4. Representative circular dichroism (CD) spectra of the most stable coiled-coil 5A1 at 20 °C. CD experiments were carried out in a 50mM PO4 (K2HPO4/KH2PO4), 100mM KCl, pH 7.0 buffer with the following conditions: oxidized disulfide-bridged 5A1 (△), reduced non-covalently-linked 5A1 (◇) and reduced 5A1 in 50% (v/v) TFE (●). Concentration of peptides are 98 μM (ox) and 196 μM (red).

Table 6-3. Biophysical data for disulfide-bridged coiled-coil analogs

Peptide [a]	Number of Consecutive Alanines [a]	Observed *M.W.* [b]	Theoretical *M.W.* [c]	$[\theta]_{222}$ [d] (degrees·cm²·dmol⁻¹)		% α-helix [e]		$[\theta]_{222/208}$ [f]		T_m [g] (°C)	$\Delta G_{app.20}$ [h] at 20 °C (kcalmol⁻¹)
				Benign	TFE	Benign	TFE	Benign	TFE		
7A1	1	11060	11000	- 35,300	- 35,000	99	99	1.03	0.91	71.4	8.6
7A3	3	11180	10916	- 33,200	- 34,900	94	98	1.02	0.89	50.2	5.1
7A5	5	10960	10832	-31,700	- 35,300	89	99	1.02	0.92	48.0	5.0
7A7	7	10800	10738	-32,000	- 34,600	90	97	1.00	0.91	46.6	5.1
6A1	1	9660	9662	- 34,700	- 34,500	99	98	1.01	0.90	79.3	9.7
6A3	3	9760	9518	- 33,800	- 34,300	96	97	1.02	0.93	59.5	2.7
6A5	5	9500	9434	- 32,200	- 34,200	92	97	1.02	0.88	56.7	2.7
5A1	1	8300	8206	- 33,600	- 34,200	97	99	1.01	0.91	87.3	11.4
5A3	3	8240	8120	- 33,900	- 34,000	98	98	1.03	0.88	70.2	8.7

[a] Peptide nomenclature, chain length and sequences are shown in Fig 6-1.

[b] Molecular weight of two-stranded coiled-coil determined by sedimentation equilibrium experiments.

[c] Molecular weight of two-stranded coiled-coil calculated based on amino acid composition.

[d] $[\theta]_{222}$ is the mean residue molar ellipticity (degrees.cm^2.dmol^{-1}) measured at 222 nm in a 100 mM KCl, 50 mM PO_4 (K_2HPO_4/KH_2PO_4) buffer, pH 7.0 in the absence (Benign) or presence of 50% trifluoroethanol (50% TFE) (v/v). Concentration of peptides ranges from 70 to 112 uM.

[e] % α-Helix was calculated from $[\theta]_{222}$ based on an ellipticity values for 100% α-helical content derived from the equation $X_{H}n = X_{H}\infty\ (1\text{-}k/n)$, where $X_{H}\infty$ is - 37,400, the wavelength dependent constant, *k*, is 2.5 and n is the number of residues in the helix.

[f] The ratio $[\theta]_{222/208}$ was calculated by dividing the observed molar ellipticity value at 222 nm ($[\theta]_{222}$) by the observed molar ellipticity value at 208 nm ($[\theta]_{208}$) in benign buffer and in 50% TFE (v/v).

[g] T_m is the temperature at which there is 50% decrease in fraction folded compared to the fully-folded coiled-coil as determined by circular dichroism spectroscopy

[h] $\Delta G_{app,20}$ is the change in free energy of the coiled-coil extrapolated from the thermal denaturation experiment at 20°C (See Materials and Methods).

observed (Fig. 6-4, Table 6-3). The ratio of $[\theta]_{222/208}$ is greater than 1 for fully folded coiled-coils in benign conditions and less than 1 for fully folded single-stranded α-helices in trifluoroethanol (Zhou *et al.*, 1992a; Sonnichsen *et al.*, 1992; Zhou *et al.*, 1993). In benign medium all of the coiled-coils showed this characteristic ratio of 1.0 to 1.03 (Table 6-3). In the presence of 50% TFE, the non-covalent hydrophobic interactions between helices were eliminated and thus, $[\theta]_{222/208}$ dropped to 0.88 to 0.93, representing the fully helical yet single-stranded state. All other model coiled-coils, from 5A1-5A3, 6A1-6A5, 7A1-7A7, including those with destabilizing Ala clusters, exhibited high helical content (89-99%) in benign buffer. Therefore, the presence of destabilizing clusters in our model coiled-coils did not have any significant effect on the α-helical content or affect the oligomeric state.

We confirmed the dimeric coiled-coil oligomerization state of all analogs by sedimentation equilibrium analyses. Fig. 6-5 shows the sedimentation equilibrium results from two representative model coiled-coils, 5A1 and 7A5, of different lengths and different size of destabilizing Ala clusters. These coiled-coil analogs showed an excellent correlation to the single species two-stranded disulfide-bridged coiled-coil unit (monomer, M), and no indication of higher oligomeric states (dimer, D or trimer, T) was observed. Observed and theoretical molecular weights are reported in Table 6-3. Taken together, the circular dichroism and sedimentation equilibrium data showed that all the coiled-coil analogs were in the fully-folded two-stranded α-helical state.

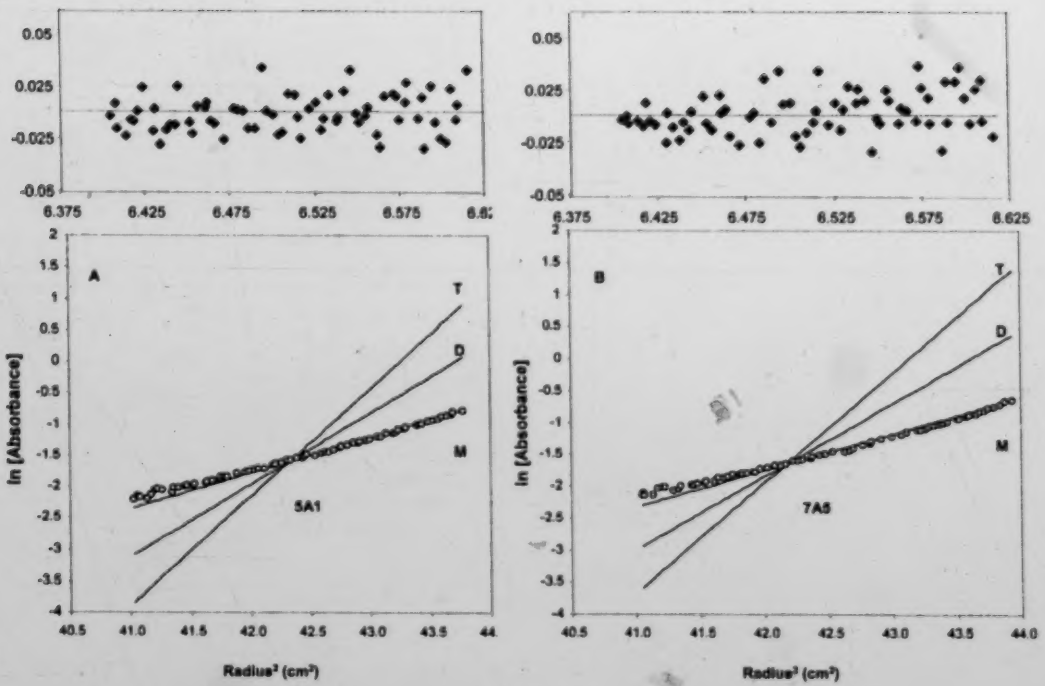


Fig. 6-5. Sedimentation equilibrium analyses of representative disulfide-bridged coiled-coils 5A1 (A) without a destabilizing cluster and 7A5 (B) with a destabilizing cluster in benign conditions (50mM PO4, 100mM KCl, pH 7.0) at 25,000 rpm: the residuals are shown above the UV absorbance data which were logarithmically best fitted to a single-species (monomeric two-stranded disulfide-bridged coiled-coil model). The global fit of the theoretical molecular weights of the monomer (M), dimer (D) and trimer (T) single species (solid lines) were plotted with the experimental data (open circles) of ln [Absorbance(274nm)] versus the distance from the center of the rotation squared. The initial peptide concentrations for 5A1 and 7A5 were 98 mM and 70 mM, respectively.

E) The effect of a three-residue alanine cluster on coiled-coil stability

Having shown that these model coiled-coils are fully-folded, we determined their thermal stabilities by circular dichroism monitored at $[\theta]_{222}$. All three coiled-coils over the temperature range of 5 to 95°C without a destabilizing cluster (5A1, 6A1, 7A1) exhibited a cooperative two-state unfolding profile with T_m of 87.3, 79.3 and 71.4 °C, respectively (Table 6-3). Coiled-coil 5A1 was the most stable of the three as expected based on our design criteria and can be attributed to a greater stability density of the five-heptad coiled-coil relative to the longer coiled-coils. The hydrophobic core density was calculated using the sum of the hydrophobic core residue stability coefficients (Wagschal *et al.,* 1999b; Tripet *et al.,* 2000; Kwok & Hodges, 2004) divided by the length in heptads.

We examined the destabilizing effect of the creation of a three alanine-residue cluster from one Leu to Ala substitution (○●○ ⇨ ○○○), comparing the stability of coiled-coils 5A1-5A3; 6A1-6A3 and 7A1-7A3, where the Leu to Ala substitution created a 3 Ala destabilizing cluster with Ala occupying consecutive ***a***, ***d***, and ***a*** heptad core positions. Despite the retention of overall helical content observed earlier, the introduction of a single destabilizing alanine cluster (○○○) from 5A1 to 5A3 drastically decreased protein stability (Fig. 6-6, panels A and D), as demonstrated by a decrease in the thermal midpoint of 17.1 °C (Table 6-4). Likewise, large drops in protein stability were observed for the creation of this 3-residue Ala destabilizing cluster (○●○ ⇨ ○○○) in models 6A1-6A3 (Fig. 6-6,

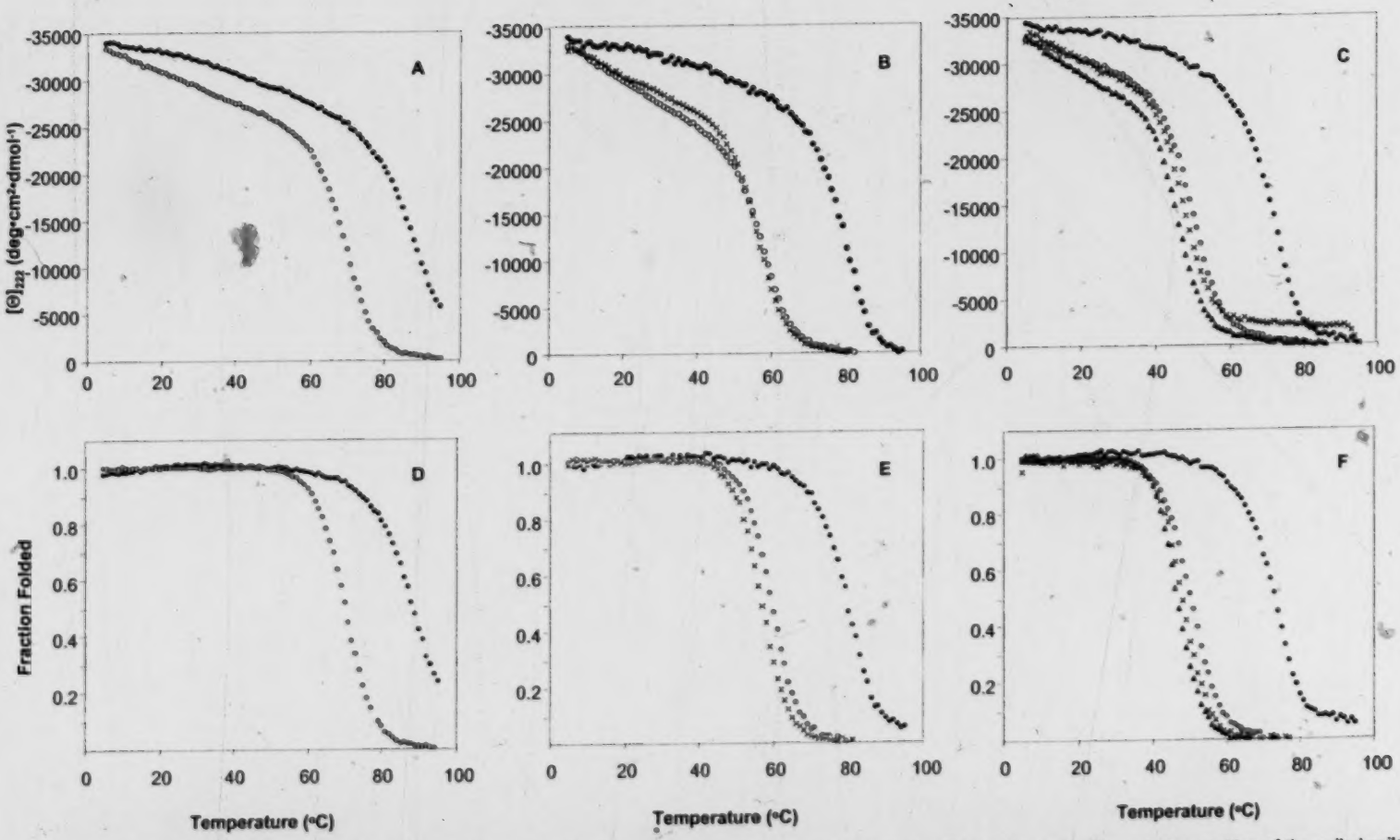


Fig. 6-6. Thermal denaturation profiles and the baseline-corrected fraction unfolding curves of the coiled-coil analogs in this study. Temperature melting of the coiled-coils was carried out in 50mM PO4 (K2HPO4/KH2PO4), 100mM KCl, pH 7.0 buffer at a starting temperature of 5 oC. Series 5 coiled-coils 5A1(●) and 5A3(○) are shown in panel A and panel D. Series 6 coiled-coils 6A1(●), 6A3(○) and 6A5(x) are shown in panel B and panel E. Series 7 coiled-coils 7A1(●), 7A3(○), 7A5(x) and 7A7(▲) are shown panel C and panel F. Concentration of peptides ranges from 70 to 98 mM. Panel D, E and F show the data corrected for pre- and post-transition baselines as described by Santoro and Bolen, 1988.

Table 6-4. Context Dependent Changes in Stability for Leu to Ala Substitutions

Location of Leu to Ala Substitution [a]	Analog Comparison	ΔT_m (°C) [b]	$\Delta\Delta G$ per Coiled-Coil (kcal/mol) [c]	$\Delta\Delta G$ per Leu to Ala Substitution (kcal/mol) [d]
○●○ → ○○○	5A1 – 5A3	17.1	2.7	1.35
	6A1 – 6A3	19.8	2.7	1.35
	7A1 – 7A3	21.2	3.5	1.75
○○○●○ → ○○○○○	6A3 – 6A5	2.8	0.0	0.00
	7A3 – 7A5	2.2	0.1	0.05
○○○○○●○ → ○○○○○○○	7A5 – 7A7	1.4	-0.1	-0.05

[a] Schematic description of the Leu to Ala substitution in the context of different numbers of destabilizing Ala residues (Fig. 6-1).
[b] ΔT_m, the change in thermal midpoint for a Leu to Ala substitution (T_m values for the individual coiled-coils are shown in Table 6-3; ΔT_m represents the decrease in T_m, for the specific Leu to Ala substitution).
[c] Relative change in the free energy of unfolding ($\Delta\Delta G$) per coiled-coil (a positive $\Delta\Delta G$ represents a decrease in stability at 20°C).
[d] Relative change in the free energy of unfolding ($\Delta\Delta G$) per Leu to Ala substitution.

panels D and E) and 7A1-7A3 (Fig. 6-6, panels C and F) with decreases in thermal midpoints of 19.8 °C and 21.2 °C, respectively (Table 6-4). This decrease in protein stability was not surprising because it corresponded to a loss of 78 $Å^2$ of non-polar surface area which was predicted to decrease stability by ~2 kcal/mol (Karplus *et al.*, 1997). We calculated this change in free energy $\Delta\Delta G$ for the creation of a 3-residue Ala destabilizing cluster to range from 2.7-3.5 kcal/mol in our models (Table 6-4). In these peptides, the Leu was in the context of a single destabilizing Ala residue on each side in the hydrophobic core, which compared favorably with a Leu-Ala destabilizing contribution of 2.6 kcal/mol reported recently by Lu and Hodges (2004) where a Leu to Ala substitution was made in the center of a three-residue hydrophobic stabilizing cluster surrounded by destabilizing residues (○○○●●●○○○ ⇨ ○○○●○●○○○).

F) The effect of leu to ala substitutions juxtaposed between a destabilizing cluster and a destabilizing residue

Coiled-coils 6A3 and 7A3 contained a Leu with an adjacent three residue alanine cluster on one side and a single destabilizing residue on the other (○○○●○). The substitution of the Leu with Ala in each case created a five residue Ala cluster (○○○●○ ⇨ ○○○○○). The question being addressed was the contribution of this leucine residue to protein stability. When we compared the stabilities of 6A3-6A5 and 7A3-7A5 (○○○●○ ⇨ ○○○○○), the loss of T_m was

marginal, with ΔT_m of 2.8 °C and 2.2 °C, respectively (Fig. 6-6, panels B, E and panels C, F; Table 6-3). This very small decrease in stability contrasted with the large decrease we observed for the creation of a three-Ala cluster (○●○ ⇨ ○○○) in the three coiled-coil models with different overall stabilities. The creation of a five-residue Ala cluster did not destabilize the coiled-coil (Table 6-4) despite the inherent identical Leu-Ala substitution.

The coiled-coil 7A5 contained a Leu with an adjacent five residue alanine cluster on one side and a single destabilizing on the other (○○○○○●○). The substitution of this Leu with Ala created a seven residue Ala cluster in the coiled-coil (○○○○○●○ ⇨ ○○○○○○○). Comparison of 7A5 and 7A7 resulted only in a decrease of 1.4 °C in T_m (Fig. 6-6, panels C and F) with no decrease in stability observed when the $\Delta\Delta G$ was calculated (Table 6-4). This observed non-linearity between protein stability and the environment surrounding the leucine residue in the hydrophobic core is critical in understanding protein folding, stability and function of coiled-coils. These results suggest that the length of a destabilizing cluster can be extremely long without significantly impacting the overall stability of the coiled-coil. For example, analogs 7A3, 7A5 and 7A7 have 3, 5 and 7 alanine clusters and essentially the same protein stability.

III) Discussion

Sequence analysis of two highly conserved and ubiquitous proteins that differ dramatically in sequence, function and length of their coiled-coils, tropomyosin (284 residues) and myosin heavy chain domain (1086 residues) showed remarkable similarity in the following ways: first, the percentage of stabilizing and destabilizing residues in the hydrophobic core (~58 and 42%, respectively); second, in both coiled-coils these residues are organized into stabilizing and destabilizing clusters and intervening regions that are well distributed along the molecules; third, the percentage of each cluster type is 60-65% stabilizing clusters and 35-40% destabilizing clusters in both proteins (Table 6-1); fourth, the percentage of intervening regions to the total number of clusters and intervening regions is 32-35% in both proteins; fifth, the ratio of destabilizing and stabilizing residues in intervening regions is close to 1:1 in both proteins.

The distribution of cluster size in stabilizing clusters and destabilizing clusters is summarized in Fig. 6-7. In myosin 88% of the stabilizing clusters are of 3, 4 or 5 consecutive residues in the hydrophobic core with a predominant cluster size of 4. In tropomyosin 100% of the stabilizing clusters are of 3, 4 or 5 residues with a predominant cluster size of 5. We have previously shown that the minimum size of a stabilizing cluster is three consecutive hydrophobic residues in the core (Lu & Hodges, 2004). In both myosin and tropomyosin the predominant size of a

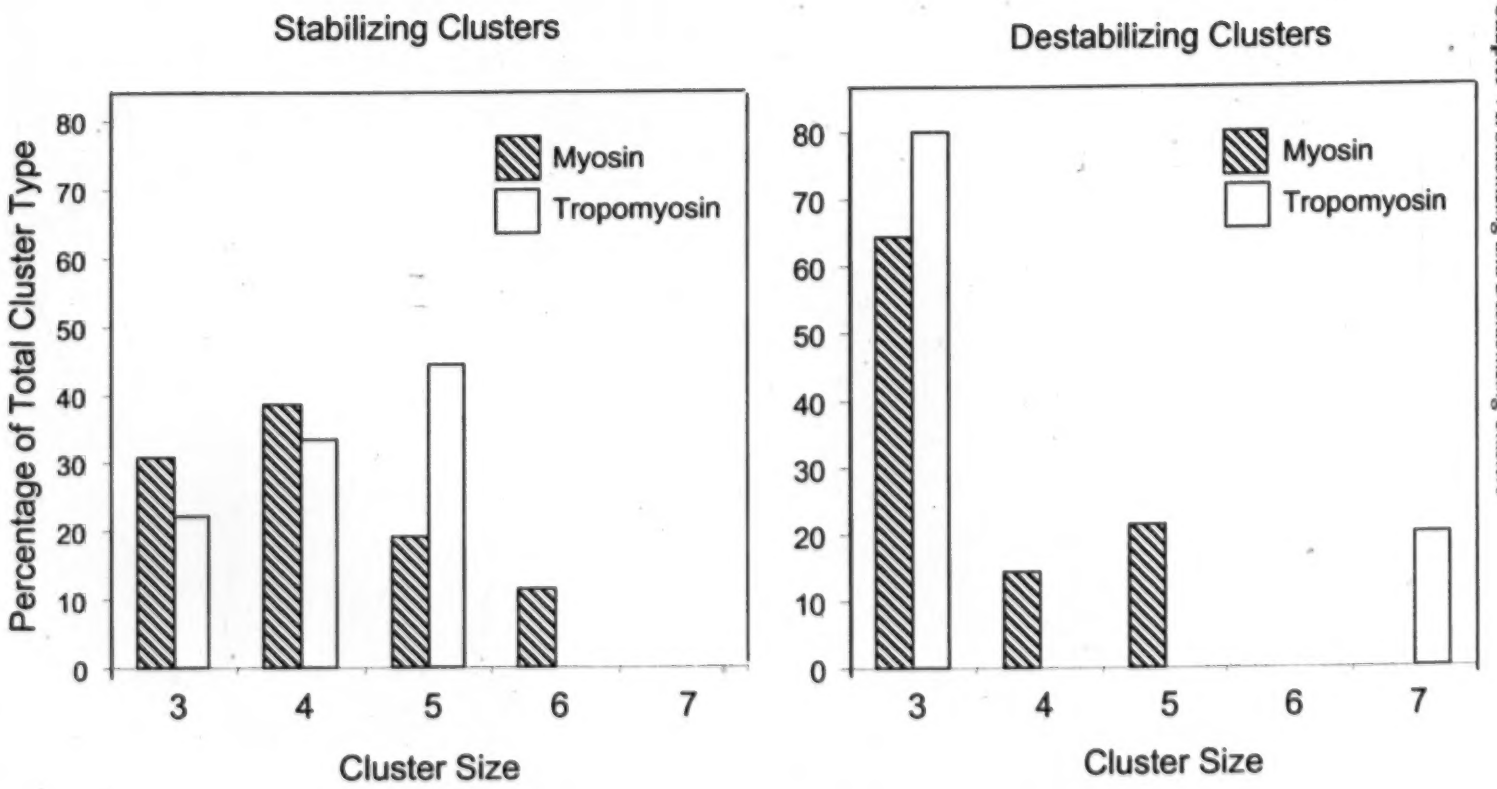


Fig. 6-7. Percentile distribution of cluster size in stabilizing and destabilizing clusters in myosin and tropomyosin. The percentages are normalized to the total number of each type of cluster in each protein molecule. Hatched rectangle: myosin; white rectangle: tropomyosin.

destabilizing cluster is 3 (Table 6-2a and 6-2b) and this agrees with our experimental results.

The dominating residue in the stabilizing clusters in both molecules is Leu (~50%) with Leu and Ile accounting for ~70% (Table 6-1). This is not unexpected since these two residues are the most hydrophobic residues of the aliphatic side-chains and can pack extremely well into the hydrophobic core of coiled-coils. The dominating residue in the destabilizing clusters in both molecules is Ala, accounting for 44% and 74% in myosin and tropomyosin, respectively (Table 6-1). This is not surprising when one considers that Ala has the highest α-helical propensity of the 20 amino acids and, though the least hydrophobic of the non-polar amino acids, Ala is the easiest to pack in the hydrophobic core because it lacks polar or charged character and still provides greater stability than the polar and charged size-chains in coiled-coils (Tripet *et al.*, 2000). This suggests that nature may want to destabilize the coiled-coil by using these destabilizing clusters and at the same time maintain the overall α-helical structure in these regions. The stabilizing and destabilizing clusters and intervening regions are well distributed along the coiled-coils and suggest an evolutionary adaptation for long coiled-coil proteins where structure integrity is maintained by the stabilizing clusters while localized regions of less stability can occur. Obviously these three types of regions control the overall protein stability, provide additional flexibility in certain regions of the coiled-coil for interactions with other molecules, control folding and assembly or allow transmission of conformational change.

In the experimental section of this study, we have clearly shown that once a destabilizing cluster of three consecutive alanine residues occurs in the hydrophobic core, extension of this cluster to seven consecutive alanine residues had little effect on α-helical structure or stability of the coiled-coil. The results shown in Fig. 6-7 show that the predominant size of a destabilizing cluster in tropomyosin and myosin is three, in agreement with the experimental results that the major destabilization of the coiled-coils is achieved with a cluster size of three. Interestingly, a cluster of 7 destabilizing residues was observed in tropomyosin (Table 6-2b). This unusual occurrence suggests a functional role to this region of the coiled-coil. Increasing the size of the destabilizing cluster could increase the flexibility of this region without affecting overall stability. Theoretical studies have suggested that local regions of instability along a coiled-coil could cause bubbling and chain separation introducing destabilizing entropy (Skolnick, 1983; Skonick & Holtzer, 1986). In contrast, short coiled-coils (<35 residues) almost always unfold from the ends because the two termini are most susceptible to end-fraying due to the lack of hydrogen bonding partners in the individual α-helices of the coiled-coil (Goodman *et al.,* 1991; Holtzer *et al.,* 1997).

Cohen and coworkers have determined the X-ray structure of a fragment of tropomyosin (residues 1 to 81, Brown *et al.,* 2001), which contained the major destabilizing cluster $K^{15}A^{18}A^{22}A^{25}K^{29}A^{32}S^{36}$ which is identical in the human and chicken sequence (Table 6-2b). The hydrophobic cluster that follows the destabilizing cluster is also identical in both species involving $L^{39}L^{43}L^{46}$ and L^{50} in

the hydrophobic core. Their analysis of the crystal structure in this region was as follows: first, this segment rich in core Ala residues (res. 22-36) was axially out-of-register by ~1.2 Å relative to canonical coiled-coils; second, the coiled-coil in this region was narrower by ~2 Å compared to the hydrophobic cluster of Leu residues that follows the destabilizing cluster; and third, there is a bend of the coiled-coil axis between the destabilizing and stabilizing cluster (in this case at ~ residue 36). The authors speculate that the structural effects of an alanine cluster on the axial register, symmetry and conformational variability of the two stranded coiled-coil appear to be important for the role of tropomyosin in the regulation of muscle contraction (Brown *et al.,* 2001). However, it should be noted that from our analysis these same destabilizing clusters occur in myosin where the myosin coiled-coil has a totally different functional role than tropomyosin. Tropomyosin has been shown to bind to actin and to move to a different position on the actin filament during muscle contraction, whereas the myosin coiled-coils are packed together in a staggered fashion to form a myosin filament. This suggests that the common occurrence of stabilizing and destabilizing clusters can have different roles in different coiled-coil proteins.

Singh and Hitchcock-DeGregori (2003) recently mutated a three-Ala cluster in tropomyosin (A_{74}, A_{78}, A_{81}) to L_{74}, V_{78}, L_{81} and showed that this change significantly stabilized tropomyosin but decreased actin affinity by >10 fold. They also made a mutant Q_{74}, N_{78}, Q_{81} which had the same stability and affinity to actin as the native protein with A_{74}, A_{78}, A_{81}. This suggests that the important determinant is

the destabilizing effect rather than the exact composition of the destabilizing residues in the destabilizing cluster.

Clearly, if we are going to understand the relationship of structure and function in coiled-coils we will have to understand the relative contribution the three regions (stabilizing and destabilizing clusters and intervening regions) make to the stability of local regions and to the overall stability of the protein.

In the native coiled-coils, tropomyosin and myosin, Leu is not only the most abundant residue in the hydrophobic core of stabilizing clusters but also the most abundant stabilizing residue in intervening regions (Table 6-1); furthermore, it is also evenly distributed between positions ***a*** and ***d***. Similarly, Ala is the most predominant residue in the hydrophobic core in destabilizing clusters (Table 6-1). Thus, in our model coiled-coils we used Leu to Ala substitutions to understand the contribution that Leu makes to protein stability in various contexts in the hydrophobic core. For example, a Leu not in a hydrophobic cluster but surrounded by one destabilizing residue on each side and adjacent to a hydrophobic cluster contributes substantially to stability (compare (●●●●○●○● ⇨ ●●●●○○○●). The ΔΔG values for the Leu to Ala substitution contributed 2.7-3.5 kcal/mol per coiled-coil (Table 6-4). The occurrence of Leu in this context in myosin was significant at 11 times. Our results suggest that in this environment in the hydrophobic core, the Leu residue still contributes significantly to the overall stability of the coiled-coil. However, once the Leu residue is adjacent to a destabilizing cluster on one side and a destabilizing residue on the other (○○○●○⇨○○○○○), the Leu residue contributes

essentially nothing to stability (Table 6-4). Interestingly, this situation occurs 6 times in myosin. The question naturally arises as to why Leu residues are found in the hydrophobic core of native coiled-coils if they do not contribute to overall protein stability. This we believe is for the preservation of the diameter and structure of the coiled-coil in a particular region. Results from the earlier described tropomyosin fragment and recent crystallographic evidence on trimeric coiled-coil Lpp-56 both showed that the core diameter of the coiled-coil with all Ala core residues was much smaller compared to a core of made up of larger bulky hydrophobes such as Leu and Ile, thus illustrating the importance of packing interactions in maintaining proper coiled-coil structure (Liu *et al.,* 2002).

Previous studies by Zhou et al. (1992a) have shown that Leu to Ala substitutions in a stabilizing cluster of a 35-residue homo-two-stranded coiled-coil (all 10 positions in the hydrophobic core were occupied by Leu) resulted in a destabilization of 2.9-3.0 kcal/mol for the central hydrophobic core positions 9, 12, 16, 19 and 23; these values are in agreement with our results. Furthermore, Lu & Hodges (2004) have recently shown that a single Leu residue with a destabilizing cluster on either side actually destabilizes the coiled-coil relative to Ala (○○○●○○○⇨○○○○○○○). Interestingly, this situation is never observed in myosin or tropomyosin, suggesting that nature does not put Leu in such a context because it would end up destabilizing the coiled-coil further (by 0.4 kcal/mol) in an already destabilized region.

Overall our experimental results and previous studies (Zhou *et al.*, 1992a; Wagschal *et al.*, 1999b; Tripet *et al.*, 2000) suggest that the surrounding environment can affect the contribution of a Leu residue to stability from a low of 0.4 kcal/mol of destabilization up to a high of 3.8 kcal/mol of stabilization, depending on its context in the hydrophobic core. This type of information is critical for the understanding of the variation in regional stability throughout a coiled-coil and the contribution of each region to overall stability, function and folding.

Chapter VII

Effect of chain length on coiled-coil stability: decreasing stability with increasing chain length

Stanley C. Kwok and Robert S. Hodges

A version of this chapter will been submitted to Biopolymers

I) Introduction

The two-stranded α-helical coiled-coil motif has a strikingly simple architecture that consists of two α-helices wound around one another in an either parallel or antiparallel fashion resulting in a left-handed supercoil (Crick, 1953; Hodges, 1996; Burkhard *et al.,* 2001). In nature, the coiled-coil motif is widely observed in more than 300 proteins involved in various cellular processes, for example, DNA-binding and gene regulation (Busch & Sassone-Corsi, 1990), muscle contraction (Singh & Hitchcock-DeGregori, 2003), intracellular vesicular transport (Sheetz, 1999; Barr, 2000), and viral entry into a host cell (Gallo *et al.,* 2003; Watanabe *et al.,* 2000). The coiled-coil domains of these proteins vary widely in length, from the short, less than 35-residues in DNA binding domains (Alber, 1992), the intermediate lengths, e.g. the 284-residue muscle protein tropomyosin (Perry, 2001), to the greater than 1000-residue myosin coiled-coil rods of muscle proteins

(Atkinson & Stewart, 1991), and this variance in size illustrates the diversity in function of coiled-coils in nature despite the simple architecture.

Manipulation of sequence in native and in *de novo*-designed coiled-coils and α-helical peptides has been employed extensively to study the interactions involved in protein folding and stability, for example, the characterization of helical propensities (Zhou *et al.*, 1993, 1994; O'Neil & Degrado, 1990; Chakrabartty *et al.*, 1991; Gans *et al.*, 1991; Li & Deber, 1994), the determination of strand association and dissociation constants (De Crescenzo *et al.*, 2003), and the quantitative contributions of hydrophobic (Wagschal *et al.*, 1999a,b; Tripet *et al.*, 2000), ionic (Kohn *et al.*, 1997, 1998; McClain *et al.*, 2001) and other non-covalent interactions (Kwok *et al.*, 2002; Kwok & Hodges, 2003; Presta & Rose, 1988; Harbury *et al.*, 1994; Tripet & Hodges, 2002). Also, the simplicity in coiled-coil secondary structure and its two-state unfolding make it an ideal model to study basic principles of protein folding. The helix-coil transition theory was first studied by Schellman in 1955 and has been extensively studied by both chemists and biologists in the past decades (Flory, 1969; Scholtz & Baldwin, 1992; Doig, 2002; Qian, 1994). Despite the extension of the helix-coil transition theories of either the Lifson-Roig (LR) or the Zimm-Bragg (ZB) theories to predict the folding and the stabilities of natural and synthetic helices (Holtzer & Holtzer, 1990; Scheraga *et al.*, 2002), relatively little experimental data on the effect of chain length on coiled-coil stability is available (Doig, 2002; Qian, 1994; Holtzer & Skolnick, 1988; Su *et al.*, 1994; Litowski & Hodges, 2001; Fairman *et al.*, 1995). In fact, the few studies that have examined the

effect of chain length on the stability of coiled-coils have suggested that the helicity and stability of coiled-coil proteins tend to increase with increasing chain length due to an increase in favorable stabilizing contacts in a cooperative non-additive manner (Su *et al.*, 1994; Litowski & Hodges, 2001; Fairman *et al.*, 1995). Increasing chain length has the obvious stabilizing effect of increasing stabilizing contacts such as hydrophobic interactions at the helix-helix interface (Wagschal *et al.*, 1999a,b; Tripet *et al.*, 2000), the possibility of introducing stabilizing intra-chain *i* to *i*+3 or *i* to *i*+4 and inter-chain *i* to *i'*+5 ionic interactions (Kohn *et al.*, 1997, 1998; McClain *et al.*, 2001; Jelesarov *et al.*, 1998; Lyu *et al.*, 1992; Stellwagen *et al.*, 1992; Scholtz *et al.*, 1993), and favorable helix-propensity effects (Zhou *et al.*, 1993, 1994; O'Neil & Degrado, 1990; Chakrabartty *et al.*, 1991; Gans *et al.*, 1991).

Despite the existing data which shows increasing stability with increasing chain length, our hypothesis for this study is shown in Fig. 7-1, that is increasing chain length can be stabilizing, neutral or destabilizing depending upon the balance of enthalpic and entropic contributions to stability. This means that if the enthalpic stabilization is greater than entropic destabilization from chain length effects, the stability of the coiled-coil will increase with increasing chain length (Fig. 7-1 - plot 1). However, if the contributions to stability are insufficient to overcome this entropic destabilization, the stability of the coiled-coil will decrease with increasing chain length (Fig. 7-1 - plot 3). Furthermore, it is conceivable that the stabilizing and destabilizing contributions are equal, thus there will be no change in stability as the chain length is increased (Fig. 7-1 - plot 2).

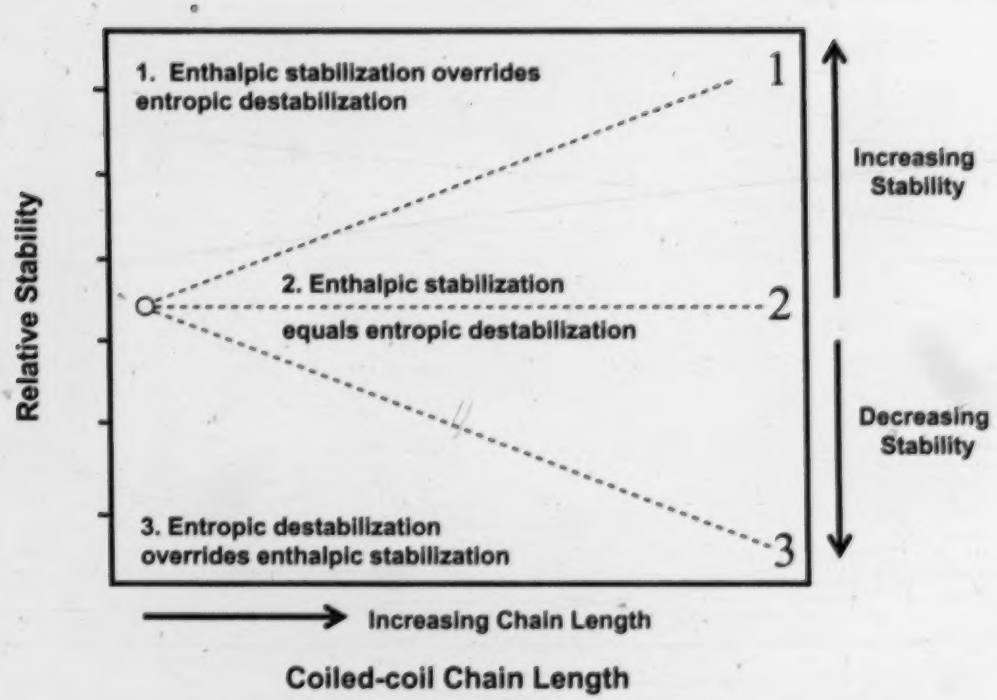


Fig. 7-1. Proposed relationship between stability and coiled-coil length depending on the sequence and overall stability of the heptad added. (1) extension of a stabilizing heptad that increases overall protein stability; (2) extension of a heptad that does not increase overall protein stability nor does it destabilize the coiled-coil stability; or (3) extension of a heptad that does not have enough stability to overcome entropic destabilization and thus overall stability decreases with chain length.

This hypothesis implies that coiled-coil stability with increasing chain length is dependent on the properties of the inserted heptad. To test this hypothesis we have designed three stable two-stranded α-helical coiled-coils that vary significantly in overall stability. We also designed a seven-residue heptad sequence for insertion (one to three times) into the central region of these three different proteins to examine the effect of chain length and whether overall stability of the coiled-coil had any influence on the chain length effect.

II) Results

A) Design of three α-helical coiled-coils with different stabilities

In the design of the proteins used in this study, we took advantage of the wealth of information provided by the de novo designed model coiled-coil proteins from previous studies (Hodges, 1996; De Crescenzo *et al.*, 2003; Zhu *et al.*, 1993; Zhou *et al.*, 1992a,b) as well as the sequence information and results on a large variety of native coiled-coil proteins (Regan & Jackson, 2003; Micklatcher & Chmielewski, 1999; Kohn & Hodges, 1998; Adamson *et al.*, 1993; Burkhard *et al.*, 2000; Hill *et al.*, 2000). Statistical surveys of the sequence of long native coiled-coil proteins showed that the hydrophobic core contains about 60% of large bulky hydrophobic amino acids (L, I, V, M, F and Tyr designated as ● for a stabilizing residue) and 40% of other smaller non-polar, polar or charged residues (designated by ○ for a destabilizing residue) (Kwok & Hodges, 2004). An alternating occurrence of clusters of stabilizing and destabilizing residues in the hydrophobic

core of coiled-coils is perhaps an evolutionary adaptation to have both regions of high stability for maintenance of structure integrity and regions of lower stability responsible for function and interactions with other protein molecules. Thus, the general approach for our design is to use heptad repeats that are representative of the hydrophobic core of long coiled-coils. There are three general types of heptad sequences that vary in the hydrophobic core positions ***a*** and ***d*** as follows: a heptad with two stabilizing hydrophobic core residues (denoted ●●); a heptad with one destabilizing residue and one stabilizing residue (denoted ○● or ●○) and a heptad with two destabilizing residues (denoted ○○). Interestingly, when you examine the >1000 residue coiled-coil myosin these three types of heptads are found 30%, 56% and 14% of the time, respectively. Thus we used the α-heptad (●●) to nucleate the coiled-coil fold; the β-heptad (○●), the most representative heptad found in native long coiled-coils to study the chain length effect and the γ-heptad (○○) to modulate overall initial stability of the coiled-coils.

Three versions of the coiled-coil proteins (A, B and C) were synthesized where the N-terminal two-heptads and the C-terminal heptad and Gly-Gly-Cys-Tyr linker were identical in the three models (Fig. 7-2). This design provided two stabilizing hydrophobic clusters (Kwok & Hodges, 2003), one at each end of the coiled-coil. The N-terminal hydrophobic cluster consisted of four consecutive large hydrophobes (Ile and Leu) in positions ***a***, ***d***, ***a*** and ***d***, respectively, and the C-terminal hydrophobic cluster consisted of three consecutive large hydrophobes, Leu, Ile and Leu in the ***d***, ***a*** and ***d***, positions respectively (Fig. 7-3).

Peptide Name	Heptad Code	Amino Acid Sequences
		gabcdef
A5	$\alpha_2\beta_2\alpha$	Ac-EIEALKA-EIEALKA-(KAEALEG)$_2$-EIEALKA-GGCY-am
A6	$\alpha_2\beta_3\alpha$	Ac-EIEALKA-EIEALKA-(KAEALEG)$_3$-EIEALKA-GGCY-am
A7	$\alpha_2\beta_4\alpha$	Ac-EIEALKA-EIEALKA-(KAEALEG)$_4$-EIEALKA-GGCY-am
A8	$\alpha_2\beta_5\alpha$	Ac-EIEALKA-EIEALKA-(KAEALEG)$_5$-EIEALKA-GGCY-am
B5	$\alpha_2\gamma\beta\alpha$	Ac-EIEALKA-EIEALKA-(KAEAAEG)$_1$-(KAEALEG)$_1$-EIEALKA-GGCY-am
B6	$\alpha_2\gamma\beta_2\alpha$	Ac-EIEALKA-EIEALKA-(KAEAAEG)$_1$-(KAEALEG)$_2$-EIEALKA-GGCY-am
B7	$\alpha_2\gamma\beta_3\alpha$	Ac-EIEALKA-EIEALKA-(KAEAAEG)$_1$-(KAEALEG)$_3$-EIEALKA-GGCY-am
C6	$\alpha_2\gamma_2\beta\alpha$	Ac-EIEALKA-EIEALKA-(KAEAAEG)$_2$-(KAEALEG)$_1$-EIEALKA-GGCY-am
C7	$\alpha_2\gamma_2\beta_2\alpha$	Ac-EIEALKA-EIEALKA-(KAEAAEG)$_2$-(KAEALEG)$_2$-EIEALKA-GGCY-am

Peptide Name	Number of Heptads	Number of β-heptad(s)	Schematic Representation of Hydrophobic residues at ***a*** and ***d*** *positions*
			a d a d a d a d a d a d a d a
A5	5	2	
A6	6	3	
A7	7	4	
A8	8	5	
B5	5	1	
B6	6	2	
B7	7	3	
C6	6	1	
C7	7	2	

Fig. 7-2. Peptide nomenclature, sequences and schematics of the coiled-coil analogs used in this study. Top panel shows the three series of coiled-coil proteins, A5-A8, B5-B7, and C6-C7 used in this study and are denoted by heptad length (e.g. B6, six heptad coiled-coil of set B). Each analog is named by the arrangement of its constituent heptad repeats described by the heptad code. Three different heptads were used (**a**: EIEALKA; **b**: KAEALEG; **g**: KAEAAEG). The chain length effect is determined using an increasing number of heptads b from 1 to 5. The heptad repeat is denoted as ***gabcdef*** which is shown above the first heptad of the coiled-coils. The residues at positions ***a*** and ***d*** in the hydrophobic core are bold. Bottom panel shows a schematic representation of the hydrophobic residues in positions ***a*** and ***d*** where large hydrophobes are shown by dark circles (Leu, Ile) and small hydrophobes are shown by open circles (Ala). The successive heptads that were inserted into the host coiled-coils are denoted by the rectangular box. Na-acetyl, Ac; C-terminal amide, am.

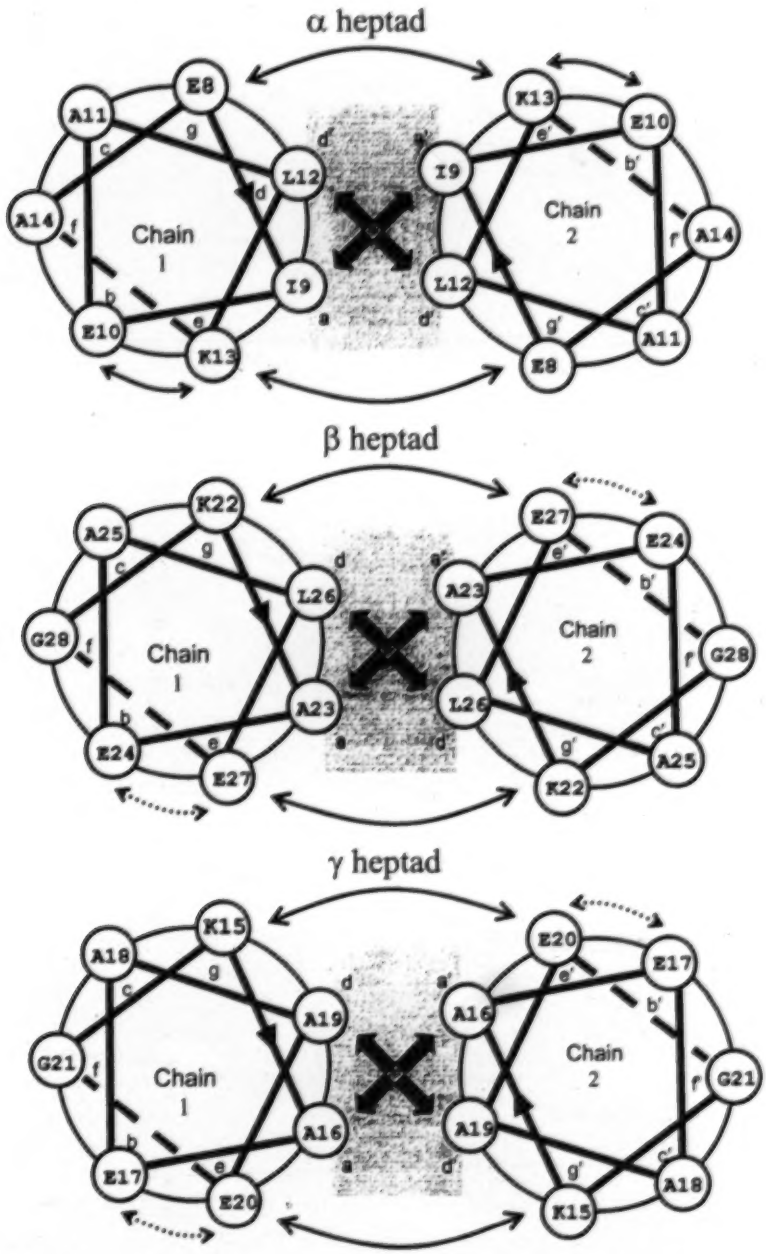


Fig. 7-3. Helical coiled-coil representations of the cross-sections of the heptads showing the different types of stabilizing interchain or intrachain electrostatic attractions (solid 2-way arrows) and destabilizing intrachain electrostatic repulsions (dashed 2-way arrows) interactions in **a**, **b**, **g** heptads. The chains are in-register and parallel, and the view proceeds from heptad position ***g*** to ***f*** (*gabcdef*) into the plane of the paper. Positions ***a*** and ***d*** and ***a′*** and ***d′*** of the heptad repeat make up the hydrophobic core (shaded rectangle). Potential intra-chain electrostatic attractions and repulsions between ***b*** and ***e*** and (***e*** and ***b***), i.e., *i, i* +3 and *i, i* +4, and potential inter-chain electrostatic attractions across the hydrophobic interface between ***g*** and ***e′*** and ***e*** and ***g′*** residues (*i, i′* +5) are denoted.

The heptad sequence used at both ends of the coiled-coil is denoted α-heptad (●●), **E-I-E-A-L-K-A**, and it provides two of the strongest stabilizing hydrophobic residues, Ile in position ***a***, and Leu in position ***d***, in the hydrophobic core of the coiled-coil (Fig. 7-3). Leu in position ***d*** is the most stabilizing hydrophobic residue (3.8 kcal/mol relative to Ala) (Tripet *et al.,* 2000), and is found most frequently in the heptad position ***d*** of native coiled-coil transcription factors (Alber, 1992). Ile is the most stabilizing hydrophobic residue at position ***a*** (3.9 kcal/mol relative to Ala) (Wagschal *et al.,* 1999b; Zhu *et al.,* 1993). Several attractive ionic interactions were also engineered to stabilize this helical heptad. Positions ***g*** and ***e*** are occupied by Glu and Lys residues, respectively, allowing the potential for inter-chain salt bridge formation (*i* to *i'*+5 or ***g*** to ***e'***) across the coiled-coil dimer interface further strengthening the interaction between α-helices (Fig. 7-3). These electrostatic interactions enhance coiled-coil stability by 0.4-0.5 kcal/mol (Kohn *et al.,* 1997, 1998; McClain *et al.,* 2001). In addition, intrachain salt bridge formation between Lys residues at position ***e*** and Glu residues at position ***b*** (*i* to *i*-3 or *i* to *i*+4 along the sequence) had been shown to enhance α-helix stability, contributing 0.2 kcal/mol (Kohn *et al.,* 1997; McClain *et al.,* 2001; Lyu *et al.,* 1992; Stellwagen *et al.,* 1992; Scholtz *et al.,* 1993). The α-heptad also has residues of high intrinsic helical propensities, Glu, Leu, Lys and Ile, and the presence of Ala, the residue of highest intrinsic α-helical propensity (Zhou *et al.,* 1993, 1994; O'Neil & Degrado, 1990; Chakrabartty *et al.,* 1991; Gans *et al.,* 1991), at solvent-exposed positions ***c*** and ***f*** to assure further the desired α-helical conformation for our model. The Gly-Gly-Cys-

Tyr linker adjacent to the C-terminal α-heptad in each polypeptide chain allows a flexible linker for the formation of the disulfide bridge to eliminate the monomer-dimer equilibrium, thus making chemical and thermal denaturation studies concentration independent (Zhou *et al.,* 1992b). Also, the disulfide bridge ensures that the polypeptide chains of the coiled-coil are parallel and in-register. The single Tyr residue allowed the determination of protein concentration by UV spectroscopy.

These three coiled-coils must provide a range of overall stabilities and sufficient stabilization for successive β-heptad insertion (○●) for the determination of chain length effects. To modulate stabilities, three versions of the coiled-coils (A, B and C) with similar compositions but different length and combinations of different helical heptads: α-heptad (●●), β-heptad (○●) and γ-heptad (○○) were synthesized. The β-heptad **K-A-E-A-L-E-G** provides an Ala in position ***a*** and Leu in position ***d*** of the hydrophobic core of the coiled-coil, in contrast to the α-heptad **E-I-E-A-L-K-A**, which has two large non-polar residues, Ile and Leu, in the core positions (Fig. 7-3). In addition, this heptad also contributes a pair of potential interchain stabilizing electrostatic attractions (*i* to *i'*+5 or ***g*** to ***e'***) across the dimer interface. There are potential intrachain destabilizing repulsions (*i* to *i*+3, ***b*** to ***e*** and *i* to *i*+4, ***e*** to ***b***); although energetically their destabilizing contribution is relatively small, only 0.2 kcal/mol for each repulsion (Zhou *et al.,* 1993). A low helical propensity and flexible Gly residue is located at the solvent exposed position ***f*** (Fig. 3). Although Gly residue is found to be very destabilizing relative to Ala in the single-stranded helical model peptides (~1.0 kcal/mol) for intrinsic helical propensity

studies (Zhou *et al.,* 1993, 1994; O'Neil & Degrado, 1990; Chakrabartty *et al.,* 1991; Gans *et al.,* 1991), the destabilization of Gly relative to Ala in a double-stranded coiled-coil is only half of the observed value in the single-stranded helix (~0.5 kcal/mol) (Kwok *et al.,* 2002). The insensitivity of the coiled-coil to Gly-Ala substitution in the solvent exposed position is because the major determinant of coiled-coil stability is the hydrophobic interactions at ***a*** and ***d*** core positions, and the identity of the residues in the outside positions ***e***, ***g*** and especially ***b***, ***c***, and ***f*** have less contribution to overall coiled-coil stability and conformation (Kwok *et al.,* 2002). Overall, compared to α-heptad, the β-heptad has less stabilizing interactions.

The third heptad, γ-heptad, **K-A-E-A-A-E-G**, contains Ala in both positions ***a*** and ***d*** of the hydrophobic core of the coiled-coil (OO); thus, it is significantly less hydrophobic in the core but yet retains the overall 3-4 hydrophobic repeat of a coiled-coil (Fig. 7-3). Otherwise the γ-heptad is identical to the β-heptad and contributes a pair of potential stabilizing interchain salt-bridge attractions (*i* to *i'*+5 or ***g*** to ***e'***) across the dimer interface, and the potential destabilizing repulsions (*i* to *i*+3, ***b*** to ***e*** and *i* to *i*+4, ***e*** to ***b***). Similarly, a low helical propensity but flexible Gly residue is located at the solvent exposed position ***f*** (Fig. 7-3).

The sequences and compositions of these three coiled-coils are designated in Fig. 7-2, and the heptad code denotes the arrangement of the combinations of α-, β-, and γ-heptads. The peptide nomenclature was used to denote the three model types by the letter A, B and C in which the number that immediately follows denotes

the number of heptads in each polypeptide chain (e.g. A5 denotes model A with 5 heptads).

B) Structural and Biophysical Characterization of Coiled-Coils (A5, B5, C6)

For the coiled-coils to be appropriate models to allow the study of chain length effect, proteins A5, B5 and C6 must exhibit a fully folded, two-stranded α-helical coiled-coil with reasonable stabilities when the disulfide bridge is formed. We employed circular dichroism spectroscopy, a sensitive probe of helical secondary structure to quantify the amount of helical structure in coiled-coil A5. The oxidized disulfide-bridged two-stranded coiled-coil A5 peptide exhibited strong helicity (Chen *et al.,* 1974), ~ 89%, in benign medium (Fig. 7-4). In the presence of helix-promoting solvent 50% TFE, no significant helical induction was observed (Table 7-1). The ellipticity $[\theta]_{208/222}$ ratio of these proteins in benign medium is generally >1.0 (Table 7-1) and distinguishes the fully folded two-stranded α-helical coiled-coil from the single-stranded α-helical conformation when the coiled-coils are placed in 50% TFE, where the ratio is less than 1.0 (Table 7-1). Even in the reduced state, this peptide is essentially fully folded with a strong $[\theta]_{208/222}$ ratio indicative of dimeric coiled-coil conformation.

We further characterized the oligomerization and aggregation state of these proteins using sedimentation equilibrium experiments by analytical ultracentrifuge

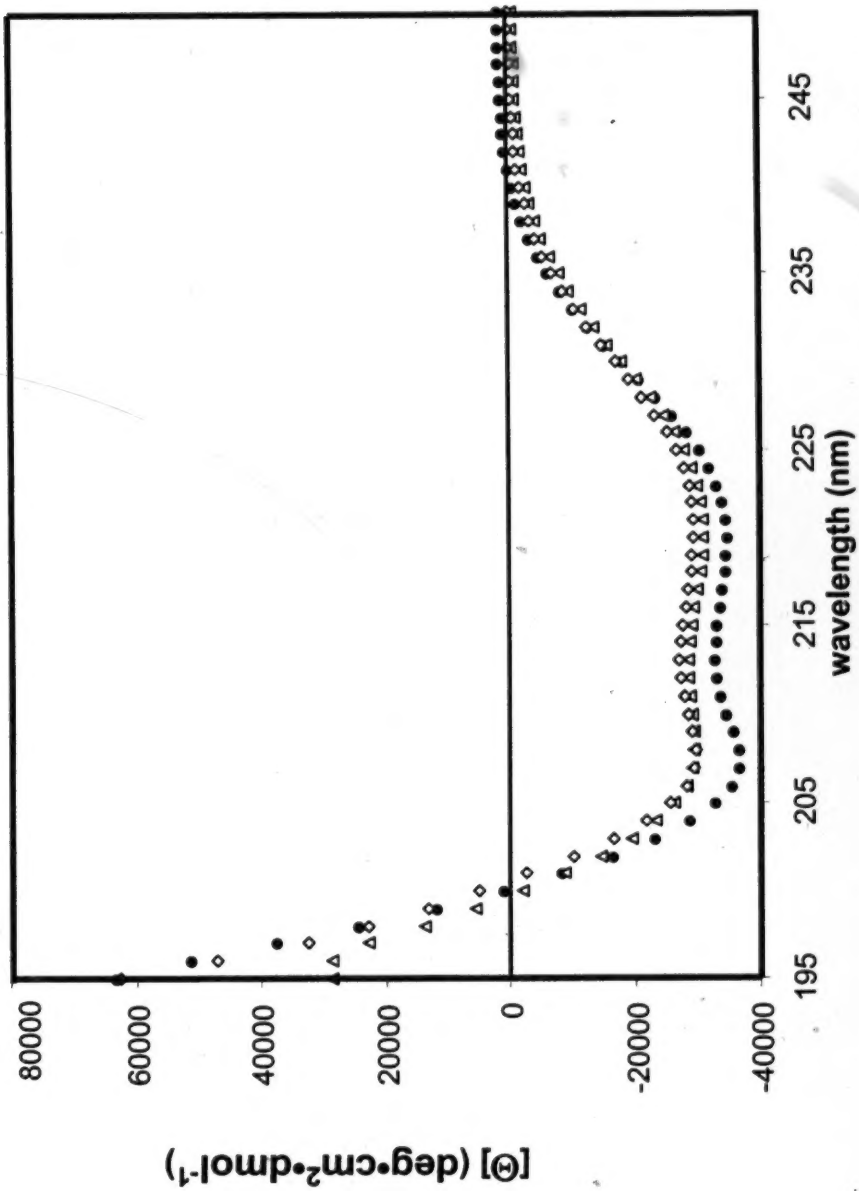


Fig. 7-4. Representative circular dichroism (CD) spectra of the most stable coiled-coil host A5 at 20 °C (Fig. 3). CD experiments were carried out in a 50mM PO4 (K_2HPO_4/KH_2PO_4), 100mM KCl, pH 7.0 buffer with the following conditions: oxidized disulfide-bridged A5 (△), reduced non-covalently-linked A5 (◇) and reduced A5 in 50% (v/v) TFE (•). Concentration of peptides ranges from 70 to 98 μM.

Table 7-1. Biophysical data for disulfide-bridged coiled-coil analogs with β-heptad inserts

Peptide [a]	Number of Heptads [a]	Number of β-heptads [a]	$[\theta]_{222}$ [b] (degrees•cm²•dmol⁻¹) Benign	$[\theta]_{222}$ [b] (degrees•cm²•dmol⁻¹) TFE	% α-helix [c] Benign	% α-helix [c] TFE	$[\theta]_{222/208}$ [d] Benign	$[\theta]_{222/208}$ [d] TFE	Observed MW [e]	Theoretical MW [f]
A5	5	2	- 30,800	- 34,200	89	99	1.01	0.93	8300	8206
A6	6	3	- 31,700	- 34,500	90	98	1.01	0.90	9660	9602
A7	7	4	- 31,300	- 33,700	88	95	1.03	0.91	11060	11000
A8	8	5	- 30,300	- 35,100	85	98	1.05	0.89	12540	12398
B5	5	1	- 30,800	- 34,000	88	98	1.03	0.88	8240	8120
B6	6	2	- 31,100	- 34,300	88	97	1.02	0.93	9760	9518
B7	7	3	- 31,000	- 34,900	87	98	1.02	0.89	11180	10916
C6	6	1	- 29,300	- 34,200	83	97	0.99	0.88	9500	9434
C7	7	2	- 28,700	- 35,300	81	99	0.97	0.92	10960	10832

[a] Peptide nomenclature, chain length and sequences are shown in Fig 7-2.

[b] $[\theta]_{222}$ is the mean residue molar ellipticity (degrees·cm^2·dmol^{-1}) measured at 222 nm in a 100 mM KCl, 50 mM PO_4 (K_2HPO_4/KH_2PO_4) buffer, pH 7.0 in the absence (Benign) or presence of 50% trifluoroethanol (50% TFE) (v/v). Concentration of peptides ranges from 70 to 98 uM. The uncertainty in the molar ellipticity values is +/- 300.

[c] % α-Helix was calculated from $[\theta]_{222}$ based on ellipticity values of -34,700, -35,200, -35,500 and -35,700 for polypeptide chains of 35, 42, 49 and 56 residues for 100% α-helical content derived from the equation $X_H^n = X_H^\infty (1-k/n)$, where X_H^∞ is - 37,400, the wavelength dependent constant, k, is 2.5.

[d] The ratio $[\theta]_{222/208}$ was calculated by dividing the observed molar ellipticity value at 222 nm ($[\theta]_{222}$) by the observed molar ellipticity value at 208 nm ($[\theta]_{208}$) in benign buffer and in 50% TFE (v/v).

[e] Observed molecular weight of two-stranded coiled-coil determined by sedimentation equilibrium.

[f] Calculated molecular weight of two-stranded coiled-coil based on amino acid composition.

Using global analyses of different protein concentrations (from 25-200μM) and rotor speeds (15,000-35,000 rpm), we determined the apparent molecular weights of our host coiled-coils (Table 7-1) and the oligomerization states that best-fitted to their sedimentation behavior. In benign buffer, these coiled-coils were predominantly single species with observed molecular weights comparable to the theoretical molecular weights of the disulfide-bridged two-stranded coiled-coil (Table 7-1). The spectroscopic and sedimentation equilibrium data together showed that all three coiled-coil proteins folded as two-stranded coiled-coils and did not assemble into higher-order oligomerization states.

Next, we characterized the stabilities of these coiled-coils by thermal denaturation monitored at $[\theta]_{222nm}$. All three coiled-coils exhibited a cooperative two-state unfolding profile with denaturation midpoints of 87.0°C, 70.2°C and 54.2°C for A5, B5 and C6, respectively (Table 7-2). Not surprisingly, coiled-coil A5 was the most stable of the three proteins because of its relatively short chain length and the presence of strong hydrophobic interactions in the hydrophobic core (3 pairs of Ile-Ile interactions in positions ***a*** and 5 pairs of Leu-Leu interactions in positions ***d*** of the hydrophobic core, Table 7-2). Demonstrating the effect of decreasing hydrophobicity in the protein core, coiled-coil B5 was less stable than A5 by 17°C because there was an Ala substitution in place of Leu in position ***d*** to generate the γ-heptad (Fig. 7-3), and this removal of the stabilizing hydrophobic side-chain, as described earlier, contributed to the decrease in stability in B5. Coiled-coil C6 was less stable than B5 by 16°C because there are two γ-heptads in its central region

Table 7-2. Stability data for disulfide-bridged coiled-coil analogs with β-heptad inserts

Peptide [a]	Number of hydrophobes in core		T_m [b] (°C)	$[Urea]_{1/2}$ [c] (M)	m [d] (kcal•mol^{-1} •M^{-1})	ΔΔG [e] (kcal•mol^{-1})	Stability [f] Density (kcal•mol^{-1})
	Leu & Ile [a]	Ala [a]					
A5	8	2	87.3	7.80	0.51	-	6.14
A6	9	3	79.3	5.55	0.88	1.55	5.75
A7	10	4	71.4	3.85	1.25	3.45	5.47
A8	11	5	63.6	2.65	1.30	4.65	5.26
B5	7	3	70.2	6.60	0.73	-	5.38
B6	8	4	59.5	4.65	0.81	1.50	5.12
B7	9	5	50.2	2.95	1.30	3.70	4.93
C6	7	5	54.2	4.60	0.90	-	4.48
C7	8	6	45.2	3.05	1.08	1.55	4.39

[a] Peptide nomenclature, chain length and sequences are shown in Fig 7-2.

[b] T_m (CD) is the temperature at which there is 50% decrease in molar ellipticity $[\theta]_{222}$ compared to the fully-folded peptide as determined by circular dichroism spectroscopy

[c] $[Urea]_{1/2}$ (CD) is the denaturation midpoint of the two-state unfolding of an α-helical coiled-coil to a random coil, i.e. the urea concentration (M) required to achieve a 50% decrease in molar ellipticity $[\theta]_{222}$ with a fully-folded coiled-coil at 20°C.

[d] *m* is the slope term associated with urea denaturation as described in Chapter II.

[e] $\Delta\Delta G_u$ Is the free energy difference contributed by adding chain length compared to the shortest analog in each series (A5, B5 or C6).

[f] The stability density was a measurement of overall hydrophobic stability divided by the number of heptads. Only the contributions of hydrophobic core residues at positions "a" and "d" were used. Ile at "a" = 3.9 kcal/mol; Leu at "a" = 3.5 kcal/mol; Ile at "d" = 3.0 kcal/mol; Leu at "d" = 3.8 kcal/mol. Ala is zero and neutral in this measurement. This gives a relative contribution to stability of the α-heptad of 8.7 kcal/mol, β-heptad of 3.8 kcal/mol and γ-heptad of 0.0 kcal/mol.

compared to one in B5, which corresponds to a string of 5 consecutive alanine residues in the core of C6 compared to just 3 consecutive alanine residues in B5. A destabilizing cluster of consecutive weakly hydrophobic alanine residues in the coiled-coil core may allow for better water penetration into the heptad ***a*** and ***d*** positions, and thus decrease the overall protein stability. Taken together, these three coiled-coils are stable scaffolds that allow for heptad insertion for the investigation of chain length effect in the context of different overall protein stabilities.

C) Successive insertion of helical heptads into the coiled-coil host A5

Having shown that our model coiled-coils are two-stranded and stable, we chose the β-heptad (Fig. 7-3) for insertion into the central region of each of the coiled-coils (type A, B and C) to increase their chain lengths. As mentioned earlier, this heptad K-A-E-A-L-E-G maintains the continuous 3-4 hydrophobic repeat of the models by placing an Ala and a Leu in the hydrophobic core positions ***a*** and ***d***, respectively. The Leu-Leu pair in the ***d*** position of the coiled-coil has been shown to contribute the most to protein stability, with an increase of 3.8 kcal/mol when compared to an Ala-Ala pair (Tripet *et al.,* 2000). Also, mentioned earlier, another stabilizing feature of this β-heptad is the two Lys to Glu electrostatic attractions ***g*** to ***e'*** (*i* to *i'*+5) that further buries the hydrophobic core away from the aqueous environment. However, this heptad sequence also introduces a pair of destabilizing intrachain electrostatic repulsions ***b*** to ***e*** (*i* to *i*+3) and ***e*** to ***b*** (*i* to *i*+4) but this destabilizing contribution has been shown to be less than <0.2 kcal/mol per

interaction (Zhou *et al.,* 1993). We chose the central region for this chain length insertion because this region should be the most sensitive to substitution (Litowskie & Hodges, 2002) because the heptads located at the ends are able to fray, making them less susceptible to chain length effects (Holtzer *et al.,* 1997).

In the case of A5, we made three chain length coiled-coil analogs, A6, A7 and A8, denoting one, two and three heptad insertions, respectively, into our most stable coiled-coil A5 (Fig. 7-2). Circular dichroism was again used to quantify the amount of helical secondary structure. At room temperature in benign condition, all three analogs were essentially fully folded α-helical coiled-coils (>85% helicity), with no significant difference relative to the A5 host. Even in 50% TFE, a helix inducing environment, no significant increase in helicity was observed for any of the chain length analogs; thus, the insertion of 7, 14 and 21 residues into the A series coiled-coil did not disrupt the overall fold. We further characterized the chain length effect on the A-series analogs by thermal denaturation, and found that the stability of the analog decreased as chain length increased (Fig. 7-5, panel C). As the number of inserted β-heptads increased from one to three, the coiled-coil was progressively less stable. Indeed, there was a linear approximation relating the temperature midpoint, T_m, to coiled-coil length (Fig. 7-6). Even in the reduced states, the decrease in stability parallels the increases in chain length among the A series analogs (Fig. 7-6). The observed linear correlation of thermal midpoint versus chain length for both the reduced and the oxidized coiled-coils exhibits nearly identical slope. Therefore, the destabilizing contribution from heptad extension is a result of the intrinsic character

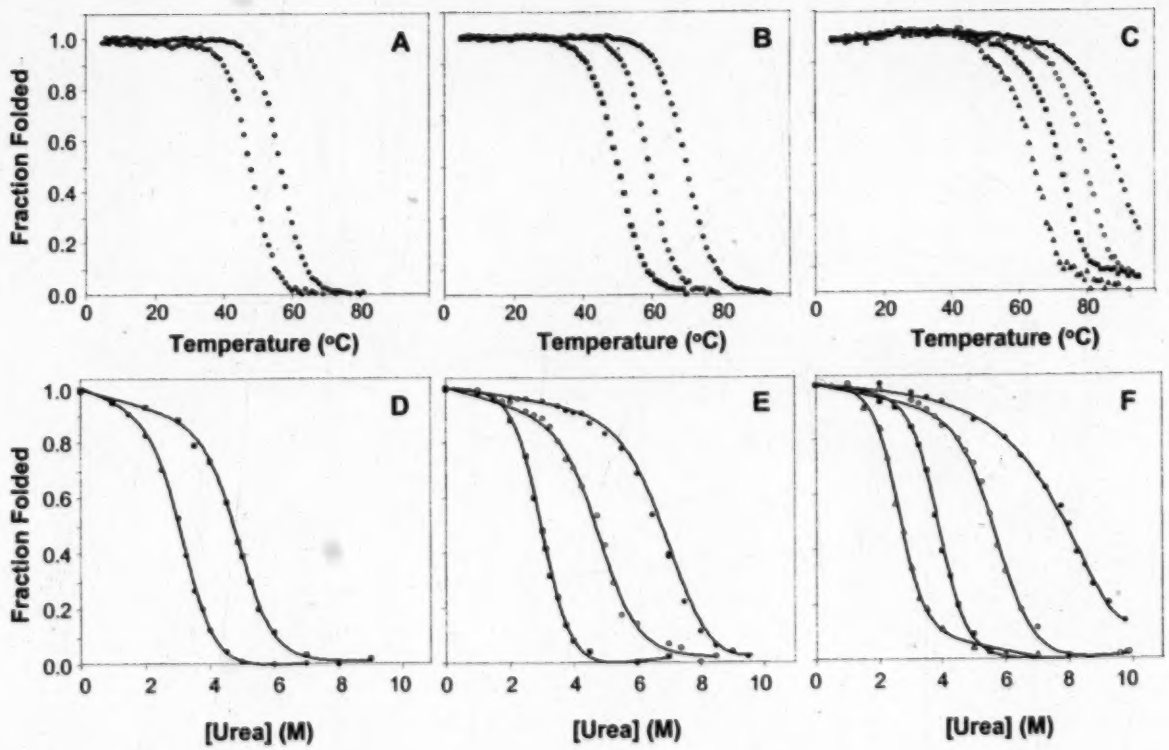


Fig. 7-5. Stability profiles of the coiled-coil analogs in this study: thermal melting (left) and urea denaturation (right). Temperature melting of the coiled-coils was carried out in 50mM PO4 (K2HPO4/KH2PO4), 100mM KCl, pH 7.0 buffer at a starting temperature of 5oC. Urea denaturation experiments were carried out at 20°C in a 50mM PO4 (K2HPO4/KH2PO4), 100mM KCl, pH 7.0 buffer with increasing concentrations of urea as denaturant. The denaturation dilution samples were vortexed and allowed to equilibrate overnight at room temperature. Series A coiled-coils A5(●), A6(□), A7(▲) and A8(○) are shown in the top panel. Series B coiled-coils B5(■), B6(△) and B7(◆) are shown in the middle panel. Series C coiled-coils C6(+) and C7(×) are shown in the bottom panel. Concentration of peptides ranges from 70 to 98 uM.

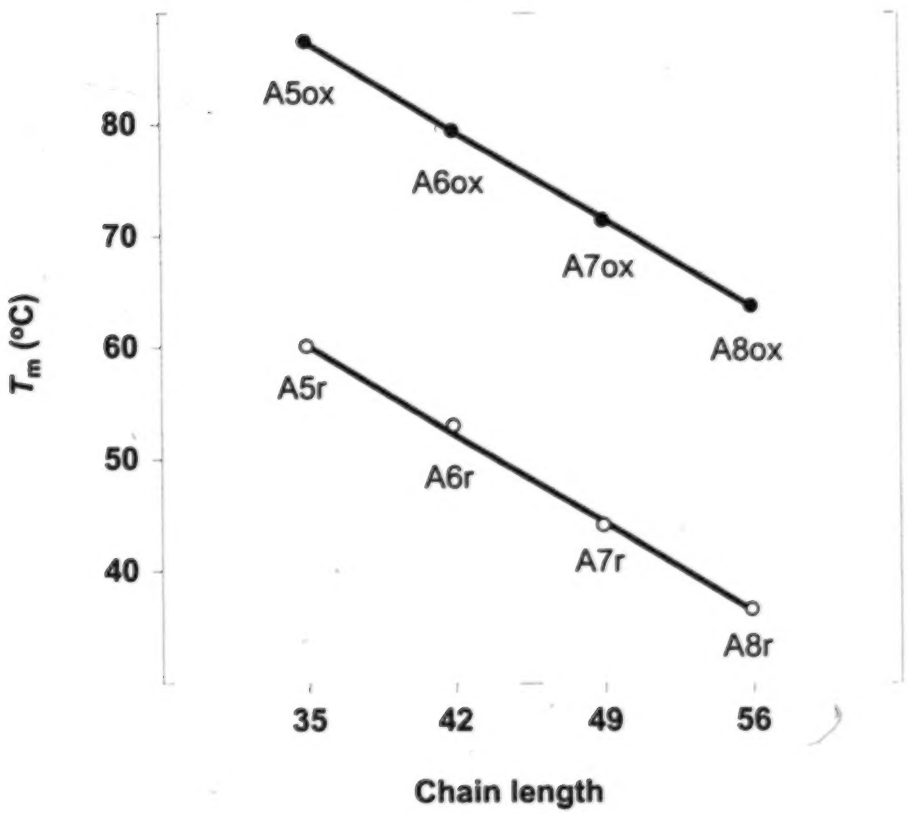


Fig. 7-6. Correlation plot of increasing coiled-coil chain length to protein thermal stability in oxidized and reduced coiled-coils of model A. Thermal midpoint (*T*m) versus the number of residues in the helix. Peptides are denoted by their heptad length and oxidized and reduced peptides are denoted as (ox) and (r), respectively.

of the "inserted heptad". To obtain useful thermodynamic parameters from the comparison of this series of analogs, chemical denaturation with urea was also performed. Similar to the trend of decreasing stabilities revealed by thermal denaturation, progressive destabilization was observed in urea midpoints ($[Urea]_{1/2}$) when the chain length was extended; however, the relationship seemed curvi-linear compared to the T_m plot (Fig. 7-7).

Parallel to the drop in protein stabilities, the *m* value, the dependence of the free energy of unfolding on denaturant concentration, progressively increased (i.e. compare A5-A8, B5-B7, C6-C7) as the coiled-coil chain lengths became longer. The denaturant *m* value had been shown to correlate strongly with the relative change in accessible surface area upon unfolding compared to the folded state (Myers *et al.*, 1995). Therefore, our derived trend of increasing *m* values (from the addition of the β-heptad(s)) correlated well with increasing peptide size and the subsequent differences in accessible surface area between the folded and unfolded states. The shortest and most stable A5 coiled-coil had the highest $[Urea]_{1/2}$ (7.8 M) and T_m (87.3°C) but the shallowest slope (m=0.51). In contrast the longest A8 coiled-coil had the lowest T_m and $[Urea]_{1/2}$ of 63.6°C and 2.6 M, respectively, but the steepest slope (m=1.30). The *m* value has been implicated in the relative difference in hydrophobic surface area between the folded and the unfolded state (Pace, 1986; Shortle, 1989) thus, although it did not affect helicity in benign media at room temperature, perhaps an increase in chain length may be exerting effects on the unfolded state.

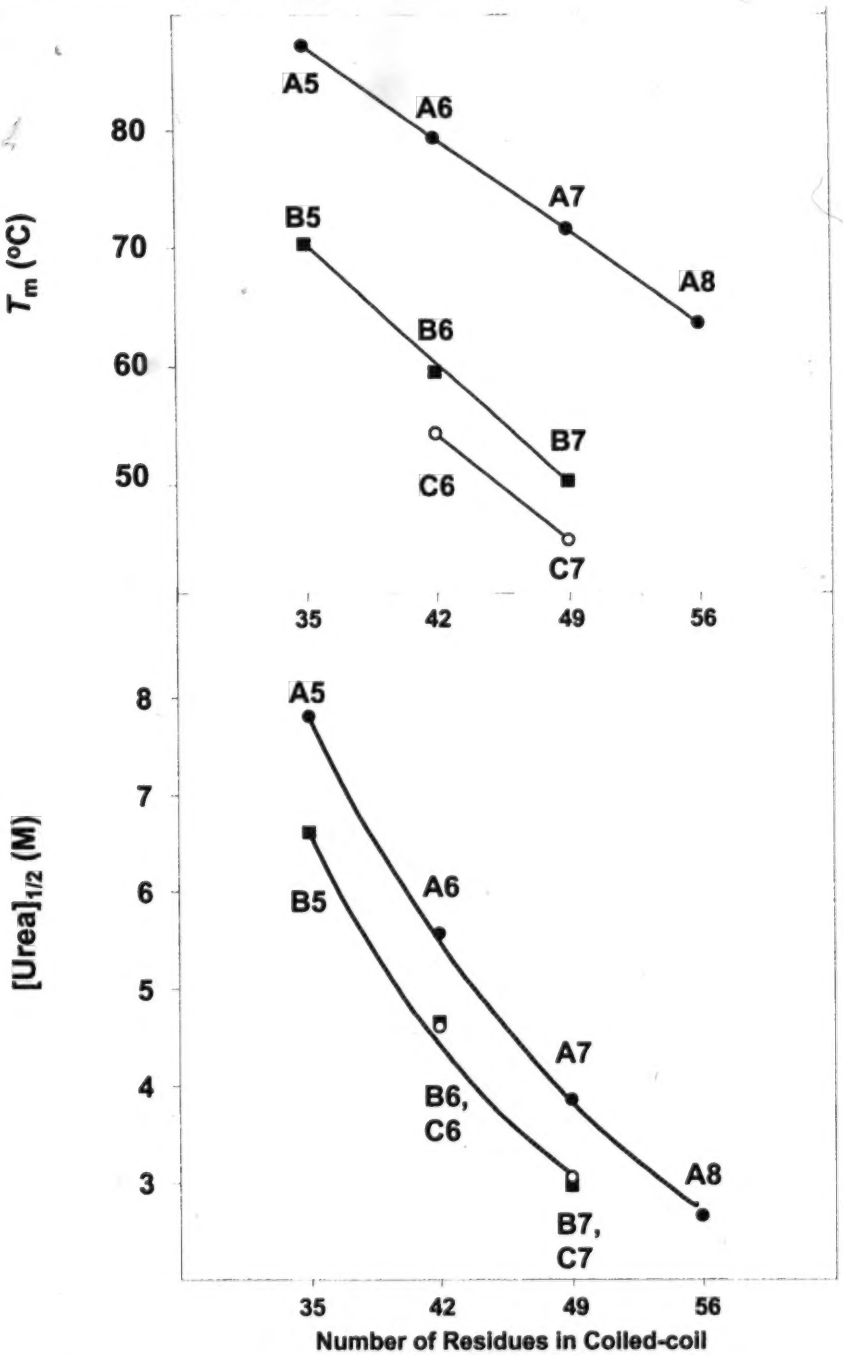


Fig. 7-8. Correlation plots of increasing coiled-coil chain length to protein thermal and urea stabilities in the three different series of oxidized coiled-coils of models A(◆), B(■) and C(○). Oxidized peptides are denoted by their heptad length. The linear plot of thermal midpoint (*T*m) versus the number of residues is shown in the top panel and the curvi-linear plot of urea midpoint ([Urea]1/2) versus the number of residues is shown in the bottom panel.

D) Helical heptad insertions into models B5 and C6

To examine if the context-independence effect of chain length insertion on overall protein conformation and stability extends to different coiled-coils, the β-heptad was similarly inserted into the central region of B5 and C6 coiled-coils. Again, each successive insertion did not disrupt the overall helical structure of the disulfide-bridged analogs; in addition no significant induction of helical structure was observed in 50% TFE. As chain length increased, the coiled-coil became progressively less stable (Fig. 7-5, Table 7-2). The slopes between chain length and the destabilization in T_m and $[Urea]_{1/2}$ in the three model coiled-coils were similar with each model, showing excellent correlations (Fig. 7-7). In addition, the comparison of T_m and $[Urea]_{1/2}$ as accurate measurements of protein stability for models A, B and C coiled-coils correlates linearly with a small deviation (Fig. 7-8). The decrease in protein stability is largely attributed to the property of the inserted heptad. Although subtle context-dependence effects may cause deviations in the slope between the three peptide series (Fig. 7-8), the destabilizing effect of chain length increase is most likely caused by the insertion of one small Ala residue and only one large hydrophobe (Leu) in the core, which is insufficient to maintain coiled-coil stability. As chain length increases, the smaller Ala, though still a non-polar residue, does not have enough non-polar surface area to maintain stability; thus, the protein decreases in stability as length increases. Interestingly, the effect of chain

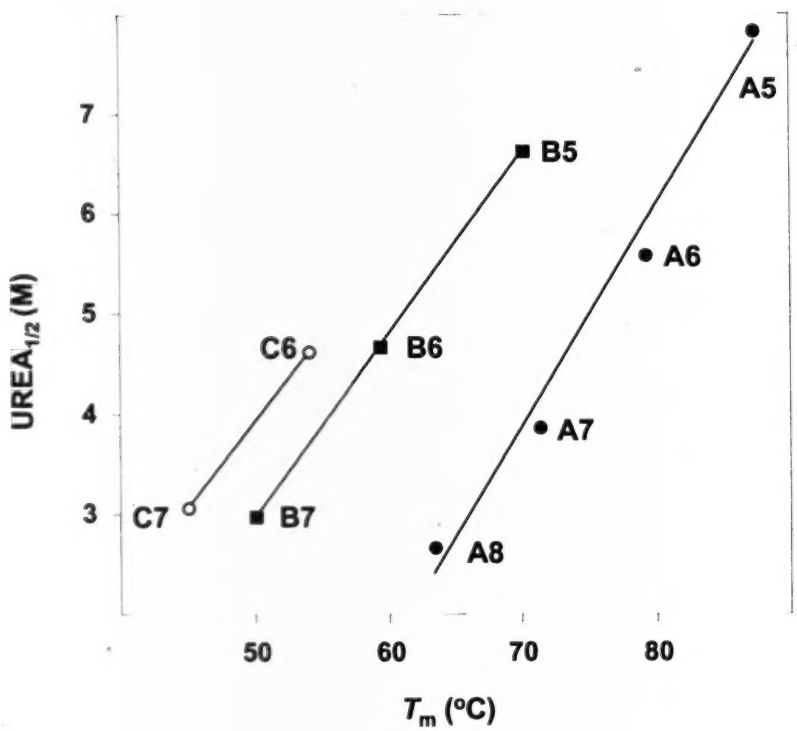


Fig. 7-8. Comparison of protein stability from thermal melting and urea denaturations in the three series of coiled-coils: A(◆), B(■) and C(○). Oxidized peptides are denoted by their heptad length. Thermal midpoint (*T*m) and urea denaturation midpoint ([Urea]1/2) are correlated.

length extension appears to be model independent as far as the overall stability of the initial coiled-coil.

III) Discussion

Traditionally, increasing coiled-coil chain length has been synonymous with increasing protein stability. Intuitively, increasing the chain length of a coiled-coil while maintaining a continuous 3-4 hydrophobic repeat increases the number of favorable stabilizing contacts and thus should enhance stability. Both Su *et al*, 1994 and Litowski and Hodges, 2001 have clearly demonstrated the stabilization of coiled-coil formation with increasing chain length up to 35-residues per polypeptide chain. In fact, there was a threshold length of 2½ to 3 heptads required for the formation of stable two-stranded α-helical coiled-coils in the absence of an interchain disulfide bridge between the two chains. In the *de novo* designed coiled-coils of Su et al, 1994, the hydrophobic core consisted of Ile at position ***a*** and Leu at position ***d***, the most stabilizing hydrophobic residues at these positions (Wagschal *et al.,* 1999; Tripet *et al.,* 2000) and even so five of these residues per polypeptide chain were required to form a stable coiled-coil.

The concepts of a threshold stability and the dramatic increases in stability of coiled-coils with increasing chain length (up to 5 heptads) are also supported by theoretical studies (Scholtz & Baldwin, 1992; Holtzer & Skolnick, 1988). Obviously, sufficient chain length is required to provide enough stabilizing interactions to reach a threshold stability such that enthalpic stabilization from

favorable interactions overcomes the entropy of fixing residues in the helical conformation (Scholtz & Baldwin, 1992. Holtzer and Skolnick (1988) simulated the stability of coiled-coils up to 40 heptads using the repetitive sequence LEALEGK (***abcdefg***) and predicted that such long coiled-coils would be thermally indestructible in agreement with experimental data. Their results also showed that stability increases dramatically at shorter chain lengths and the increase in stability with increasing chain length diminishes as the polypeptide chain gets longer.

A further consideration in increasing helix chain length is the "end-fraying" phenomenon where the partial and temporary unwinding of the coiled-coil occurs at the helix termini. In contrast, the helical heptads in the central location of the coiled-coil are more tightly packed; thus, they are less susceptible to "end-fraying" and therefore contribute more to protein stability (Holtzer *et al.,* 1997; Hodges *et al.,* 1990). As chain length increases, the destabilizing contribution of end-fraying per heptad is diminished, as more heptads are sequestered in the center of the coiled-coil.

In this study, there are four entropic terms to be considered in the folding of coiled coils: **1)** we introduced the disulfide bond between the two polypeptide chains thus reducing the translational and rotational entropy (Zaman *et al.,* 2002); **2)** the entropy term associated with helix formation; **3)** the entropy term associated with docking or binding of the two α-helices to generate the coiled-coil; and 4) the entropy term associated with increasing chain length. In our model peptides, we chose to insert the helical heptads into the center of these stable coiled-coil proteins, containing at each termini the identical stabilizing hydrophobic clusters (Kwok &

Hodges, 2003) and an interchain disulfide bridge (Zhou *et al.,* 1992b) at the C-terminus. These three features, the disulfide-bridge, the stabilizing clusters and the center guest site are common to all the coiled-coil analogs in this study. Therefore, the first three entropic terms listed above should be near identical, thus leaving the chain length effect as the only unaccounted for entropic term. This interpretation is supported by the fact that despite the decrease in protein stabilities in all chain length analogs, the coiled-coil motif remained folded. This implies that we have isolated the chain length entropic term from the other common entropic terms involved.

Another approach to interpreting our results is the concept of hydrophobic density. That is, the overall stability of the coiled-coil, calculated from the sum of the contributions of the hydrophobic core residues at positions ***a*** and ***d***, divided by the number of heptads in the coiled-coil to give a hydrophobic density score per heptad. We computed the average hydrophobic density per heptad for all our analogs based on the published thermodynamic parameters from host-guest studies (Wagschal *et al.,* 1999a,b; Tripet *et al.,* 2000) and found that successive extension of helical length by adding an heptad with Ala and Leu in the core resulted in decreasing hydrophobic density (Table 7-2). The ***a*** and ***d*** heptad positions at the ends of the three host coiled-coils consist of Ile and Leu, two of the most stabilizing residues in the core. Ile at position ***a*** is 3.9 kcal/mol and Leu at position ***d*** is 3.8 kcal/mol more stable than Ala which provides a total of 7.7 kcal/mol stabilization for these heptads. In contrast each heptad insertion in this study added one Leu-Leu interaction and one Ala-Ala interaction, a total of 3.8 kcal/mol stabilization per

heptad. The classic observation that extending coiled-coil length increases protein stability is probably attributed to the fact that the inserted heptads have the same stability density as the heptads in the initial protein; thus the hydrophobic density is maintained. However in the present study, as successive heptads were inserted, the inserted hydrophobic stabilization was insufficient to maintain the hydrophobic density observed in the initial coiled-coil and therefore the proteins decreased in stability. Perhaps other coiled-coil designs would allow an increase in chain length with no increase in stability, i.e., the enthalpic stabilization equals entropic destabilization as the helix is extended (Fig. 7-1 - Plot 2). Interpretation of our results suggests that the introduction of one pair of favorable hydrophobic stabilizing interactions (Leu-Leu) in the core per heptad did not provide sufficient enthalpic stabilization to overcome the destabilization entropy of helical extension and the stability of the entire molecule therefore decreased.

Clearly, the chain length effect is context-dependent (Minor & Kim, 1996) and related to the composition of the host, the inserted heptad and the location of insertion. The increase of coiled-coil chain length or the change in hydrophobic density might have profound influence on the hydrophobic contacts of the unfolded states, thus complicating the context-dependent interactions even further.

In nature, a balance of the number of regions of higher and lower stability is the determinant of protein function. The stable domains of a protein provide the necessary scaffold to maintain structure and integrity while allowing the flexible unstable regions to be involved in function. Perhaps hydrophobic clustering and

hydrophobic density are more appropriate descriptions of coiled-coil folding than overall hydrophobicity and chain length; we are currently taking advantage of these new observations and incorporating them in the coiled-coil prediction program, STABLECOIL (biomol.uchsc.edu/researchfacilities/computational/core/stablecoil) to achieve more accurate predictions of protein stability (Tripet & Hodges, 2001).

Chain length effect has significant biological relevance in protein folding pathways. For example, the cellular ribosomal machinery synthesizes protein sequentially from the N-terminus towards the C-terminus. Therefore, the *in vivo* protein folding process could begin before the full length protein chain has been assembled. Thus the elucidation of the folding propensities of different chain lengths is paramount to understanding the protein folding pathway. For example, the folding of apomyoglobin, a single-domain protein consists of all α-helices and intervening loops, is highly chain length dependent (Chow *et al.*, 2003). At relatively short lengths from the N-terminal, a predominantly non-native β-sheet conformation is present, and misfolded amyloid-like structures are the folding intermediates. As the protein chain length extends to the C-terminal, the β-strand content diminishes and the fraction of α-helix content increases. The authors postulate that the C-terminal region of this protein plays an important role by promoting helix formation and induces the co-operative packing of other shorter helices located near the N-terminus, thus guiding the secondary structures to fold into the proper tertiary conformation observed in the native protein. Further investigations into this area

will provide new insights in the complexities of coiled-coil folding, and should be applicable to protein folding in general.

Chapter VHI

Future Directions

I) Overview

The research of this thesis demonstrates the application of the two-stranded α-helical coiled-coil motif as an excellent model to study the interactions for protein folding and stabilities. Many possibilities exist for future studies using peptide technology to solve biological problems, and one example would be the on going collaborative project with Dr. Dean Edwards of University of Colorado Health Sciences Center to examine the interactions of progesterone receptor with its ligands. In addition, several areas for potential future research have already been mentioned in the preceding chapters, and only two aspects will be described here: **1)** using the structural cassette mutagenesis model to induce the conversion of α-helical structure to β-sheet conformation and 2) the application of RP-HPLC temperature profiling to examine hydrophobic clustering phenomenon.

A) Conversion to β-sheet amyloid structure

The studies in Chapter III and IV demonstrate that it is possible, under optimal conditions, to insert a short β-sheet segment (TSATLAITGLQ) from a native protein into a *de novo* designed coiled-coil host protein and maintain overall

helical structure. In contrast, it is also possible that a β-sheet with no propensity to form a helical conformation (NNARFSVSKSG) can prevent the coiled-coil host from folding. Thus this particular chimera construct becomes random-coil, but only small β-sheet content is observed. Therefore, the natural criticism of the structural cassette mutagenesis model is that the host coiled-coil with two flanking nucleating α-helices is too stable to allow formation of β-conformations.

To decrease the stability of the host coiled-coil thus reducing the potential to nucleate helical structure, we introduced a single hydrophobic substitution (Lys ⇨ Leu) into the N-terminal nucleating α-helix (Fig. 8-1). The rationale of this design is to increase the overall hydrophobicity of model (known to enhance β-sheet formation) and to disrupt the amphipathicity of the nucleating α-helices (destabilize the preferred α-helical conformation). The preliminary structural characterization results showed not only that this new construct exhibited very little helical structure, but the soluble peptide slowly aggregated with increasing β-sheet content, and eventually forming an insoluble amyloid matrix. The formation of the β-sheet amyloid was a concentration-dependent and time-dependent process that was irreversible. These preliminary results showed the significance of hydrophobicity in controlling protein folding, and further biophysical characterization of this aggregation behavior in different buffer systems, with different additives and varying amount of proteins will give us insights to the process of inter-conversion of helical structure, β-sheet structure and amyloid aggregate.

Parent Peptide
1 5 10 15 20 25 30 35 39
f g a b c d e f g a b c d e f g a b c d e f g a b c d e f
H2N—C-G-G-K-E-V-E-A-L-K-A-Q-V-N-N-A-R-F-S-V-S-K-S-G-A-E-V-E-A-L-K-K-E-V-E-A-L-K-K-amide
H2N—C-G-G-K-E-V-E-A-L-L-A-Q-V-N-N-A-R-F-S-V-S-K-S-G-A-E-V-E-A-L-K-K-E-V-E-A-L-K-K-amide
Hydrophobic analog K10L

Fig. 8-1. Sequence comparison of the parent peptide (Chapter IV) versus the hydrophobic analog (K10L) highlighting the single hydrophobic substitution in position 10. Parent peptide shows a soluble random-coil conformation. Hydrophobic analog K10L forms amyloid in benign condition.

B) Temperature profiling to characterize hydrophobic clustering

A recent development in our laboratory is the application of RP-HPLC technology to derive thermodynamics parameter describing the partitioning behavior of a sample between the mobile phase (liquid) and the solid support matrix. This new temperature profiling approach is a sensitive technique to measure "apparent" hydrophobicity and to describe dimerization behavior.

In general, the retention time of a non-associating, non-interacting protein on a RP-HPLC column generally decreases with increasing temperature, i.e., the sample elutes earlier at higher temperature. This behavior is attributed to increasing solubility of the protein (sample) in the mobile phase (organic solvent) and decreasing affinity for the hydrophobic support matrix. In contrast, the retention coiled-coils does not follow this trend (Fig. 8-2A.). The retention time of these molecules increased as temperature rose, up to a maximal, $t_{r(max)}$, and once past this threshold, the retention time decreased, resulting in a parabolic profile (Fig. 8-2B.).

We used $t_{r(max)}$ and Δt_r to describe the difference in binding interactions of coiled-coils with different hydrophobic clustering. Because of its sensitivity to small changes in the non-polar protein core, this technique was able to distinguish between the two different hydrophobic clustered peptides, P3 and P2 (described in Chapter V) in preliminary studies. The three-clustered coiled-coil P3 had a higher $t_{r(max)}$ and Δt_r when compared to P2, indicating that P3 was more retentive and thus more hydrophobic.

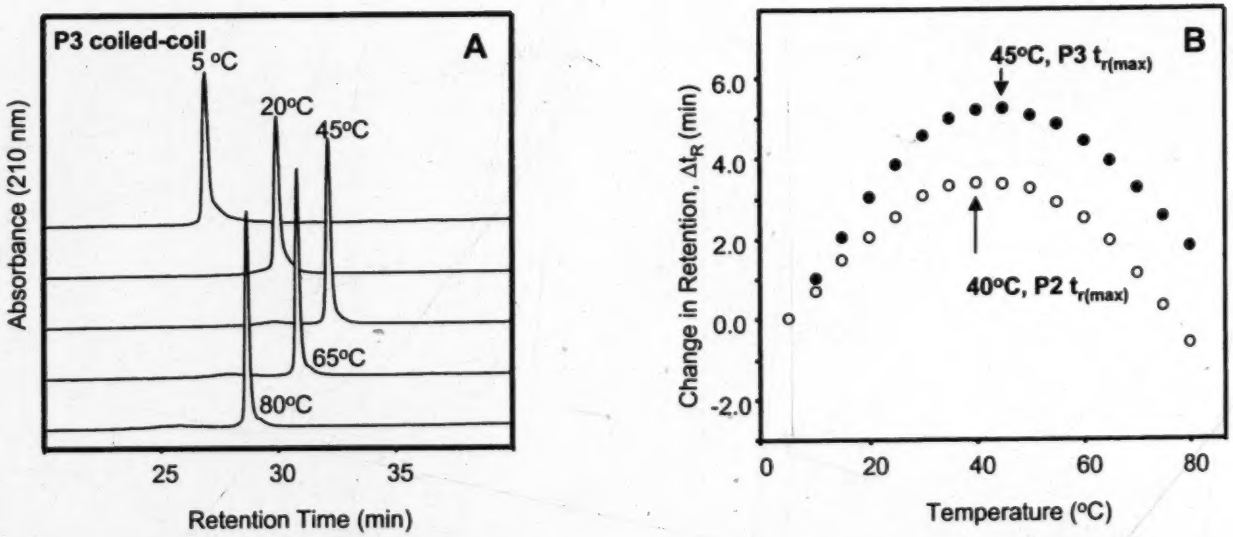


Fig. 8-2. Temperature profiling by RP-HPLC to measure stability and association of α-helices to form a two-stranded coiled-coil. A) P3 coiled-coil HPLC profiles at different temperature. B) Retention behavior comparison beween P3 and P2 coiled-coil. RP-HPLC conditions: Zorbax SB-C_8 column, linear AB gradient (1% $Bmin^{-1}$) at a flow rate of 0.25 $mLmin^{-1}$ where eluent A is 0.05% aq. TFA and eluent B is 0.05% TFA in acetonitrile.

We hypothesized that the difference in parabolic retention behaviors is due to the ability of RP-HPLC to detect the subtle structural and stability variance between these two coiled-coils. Further experiments are underway to examine other coiled-coil analogs and to derive a model to explain how coiled-coils behave differently compared to non-associating molecules in the RP-HPLC environment.

II) Impact and outlook

This thesis described the relevance of secondary structural specificity determinants and hydrophobic clustering in controlling protein conformation and stability. These phenomena represent common protein characteristics that nature utilizes to regulate regions of stability for maintenance of protein integrity and regions of instability for optimal interactions with other biomolecules. Structural biologists who study evolutionary trends of protein sequences have suggested that the emergence of destabilized protein regions may represent eukaryotic adaptations to create more diverse protein interactions. In addition to the "NN" motif described in our research, other secondary structural motifs involving di-serine and di-threonine sequences have been identified as conferring specific secondary structure.

Rather than understanding protein folding one amino acid residue at a time, these important motifs can be useful markers to understand the relationship of protein folding and stability. If each individual amino acid residue can be considered to be part of an "alphabet" code, motifs such as specific hydrophobic clustering and

secondary structural specificity determinants would represent "words" that translate into protein structure. These "words" would be key markers for structure determination in a linear protein sequence that may contain structurally ambivalent "chameleon" sequences (regions of redundancy and extraneous information). This classification scheme will be a marked step forward to understand protein folding and stability, as we continue to deconvolute and characterize sequences that are structurally and functionally important and others that are not, i.e., simplifying the protein folding problem.

Understanding protein folding is still one of the most important and difficult challenges of modern day science. The studies in this thesis stressed the importance of incorporating both the features of short-range interactions (for example, secondary structural specificity determinants) and long-range interactions (for example, hydrophobic clustering) in understanding protein folding. Both of these phenomena and the intricate context-dependent balance between them are what controls the final fold of a protein.

These investigations represent novel approaches to understand protein stability by identifying sequences which are critical for stability and function. Rather than studying protein folding one residue at a time, we described universal stabilization properties such as secondary structural specificity determinants and hydrophobic clustering that represent "words" or higher order information packets among the vast and often redundant data encoded in the primary amino acid sequence. Identifying such key sequences or motifs is essential to describe

structurally and functionally important regions in proteins and distinguishes them from the less important sequences that are structurally ambivalent. The critical examination of amino acid sequence would minimize extraneous information encoded in "chameleon" sequences and would also facilitate computer algorithms for the prediction of protein structure and stability. Not only do these examples illustrate the intricacies of context-dependent stabilization in secondary structures but, more importantly, it has universal relevance to understanding protein folding, stability and *de novo* design.

It is an optimistic outlook that these studies and other *de novo* protein research will bring new information on biological molecules, thus providing further insights into all life processes. Ultimately, the goal of a structural biologist is to understand and emulate all the activities of native proteins, and even to improve on them through knowledge-based design. The future is filled with possibilities of novel proteins such as catalysts, receptors and ligands that would have many industrial and biological applications.

BIBLIOGRAPHY

Adamson, J. G., Zhou, N. E. & Hodges, R. S. (1993). Structure, function and application of the coiled-coil protein folding motif. *Curr Opin Biotechnol* 4, 428-437.

Alber, T. (1992). Structure of the leucine zipper. *Curr Opin Genet Dev* 2, 205-210.

Albericio, F., & Carpino, L. A. (1997). Coupling reagents and activation. *Methods Enzymol* 289, 104-126.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-230.

Aqvist, J., Luecke, H., Quiocho, F. A., & Warshel, A. (1991). Dipoles localized at helix termini of proteins stabilize charges. *Proc Natl Acad Sci USA*, 88, 2026-2030.

Araya, E., Berthier, C., Kim, E., Yeung, T., Wang, X., & Helfman, D. M. (2002). Regulation of coiled-coil assembly in tropomyosins. *J Struct Biol* 137, 176-183.

Argos, P. (1987). Analysis of sequence-smilar pentapeptides in unrelated protein tertiary structures. Strategies for protein folding and a guide for site-directed mutagensis. *J. Mol. Biol.* 197, 331-348.

Atkinson, S. J. & Stewart, M. (1991). Molecular basis of myosin assembly: coiled-coil interactions and the role of charge periodicities. *J Cell Sci Suppl* 14, 7-10.

Aurora, R., & Rose, G. D. (1998). Helix capping. *Protein Sci* 7, 21-38.

Baldwin, R. L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proc Natl Acad Sci USA* 83, 8069-8072.

Barlow, D. J., & Thornton, J. M. (1983). Ion-pairs in proteins. *J Mol Biol* 168, 867-885.

Barr, F. (2000). Vesicular transport. *Essays Biochem* 36, 37-46.

Behe, M. J., Lattman, E. E., Rose, G. D. (1991). The protein-folding problem: the native fold determines packing, but does packing determine the native fold? *Proc Natl Acad Sci USA* 88, 4195-4199.

Berg, J. S., Powell, B. C., & Cheney, R. E. (2001). A millennial myosin census. *Mol Biol Cell* 12, 780-794.

Betz, S. F., Liebman, P. A. & DeGrado, W. F. (1997). De novo design of native proteins: characterization of proteins intended to fold into antiparallel, rop-like, four-helix bundles. *Biochemistry* 36, 2450-2458.

Biswas, K. M., DeVido, D. R. (2003). Evaluation of methods for measuring amino acid hydrophobicities and interactions. *J Chromatogr A* 1000, 637-655.

Blaber, M., Zhang, X. J., & Matthews, B. W. (1993). Structural basis of amino acid alpha helix propensity. *Science* 260, 1637-1640.

Blaber, M., Zhang, X. J., Lindstrom, J. D., Pepiot, S. D., Baase, W. A., Matthews, B. W. (1994). Determination of alpha-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. *J Mol Biol* 235, 600-624.

Bodanszky, M. (1998). *Peptide Chemistry*. Berlin, Springer-Verlag.

Bodanszky, M. (1993). *Peptide Chemistry*. Berlin, Springer-Verlag.

Borg, J., Jensen, M. H., Sneppen, K., and Tiann, G. (2001). Hydrogen bonds in polymer folding. *Phys. Rev. Lett.* 86, 1031-1033.

Brown, J.H., Kim, K. H., Jun, G., Greenfield, N. J., Dominguez, R., Volkmann, N., Hitchcock-DeGregori, S. E., & Cohen, C. (2001). Deciphering the design of the tropomyosin molecule. *Proc Natl Acad Sci USA* 98, 8496-8501.

Burkhard, P., Meier, M. & Lustig, A. (2000). Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci* 9, 2294-2301.

Burkhard, P., Stetefeld, J. & Strelkov, S. V. (2001). Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol* 11, 82-88.

Busch, S. J. & Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet* 6, 36-40.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., & Mornon, J.P.. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci* 53, 621-45.

Carr, C. M., & Kim, P. S. (1993). A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73, 823-832.

Chakrabartty, A., Kortemme, T., & Baldwin, R. L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* 3, 843-852.

Chakrabartty, A., Schellman, J. A. & Baldwin, R. L. (1991). Large differences in the helix propensities of alanine and glycine. *Nature* 351, 586-588.

Chana, M., Tripet, B. P., Mant, C. T., & Hodges, R. S. (2002). The role of unstructured highly charged regions on the stability and specificity of dimerization of two-stranded alpha-helical coiled-coils: analysis of the neck-hinge region of the kinesin-like motor protein Kif3A. *J Struct Biol* 137, 206-219.

Chao, H., Houston, M. E. Jr, Grothe, S., Kay, C. M., O'Connor-McCourt, M., Irvin, R. T., & Hodges, R. S. (1996). Kinetic study on the formation of a de novo designed heterodimeric coiled-coil: use of surface plasmon resonance to monitor the association and dissociation of polypeptide chains. *Biochemistry* 35, 12175-12185.

Chen, Y-H., Yang, J. T. & Chau, K. H. (1974). Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* 13, 3350-3359.

Chou, P. Y., & Fasman, G. D. (1974a). Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry* 13, 211-222.

Chou, P. Y., & Fasman, G. D. (1974b). Predictions of protein conformation. *Biochemistry* 13, 222-245.

Chou, P. Y., & Fasman, G. D. (1978). Empirical predictions of protein conformation. *Annu Rev Biochem* 47, 251-276.

Chow, C. C., Chow, C., Raghunathan, V., Huppert, T. J., Kimball, E. B. & Cavagnero, S. (2003). Chain length dependence of apomyoglobin folding: structural evolution from misfolded sheets to native helices. *Biochemistry* 42, 7090-7099.

Choy, W. Y., Shortle, D., & Kay, L. E. (2003). Side Chain Dynamics in Unfolded Protein States: an NMR based 2H spin relaxation study of Δ131Δ. *J Am Chem Soc* 125, 1748-1758.

Cohen, C. & Parry, D. A. (1990). Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. *Proteins* 7, 1-15.

Cohen, F. E. (1999). Protein misfolding and prion diseases. *J Mo Biol* 293, 313-320.

Colon, W., Elove, G. A., Wakem, L. P., Sherman, F., & Roder, H. (1996). Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. *Biochemistry* 35, 5538-5549.

Cooper, T. M. & Woody, R. W. (1990). The effect of conformation on the CD of interacting helices: a theoretical study of tropomyosin. *Biopolymers* 30, 657-676.

Creamer, T. P., & Rose, G. D. (1992). Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc Natl Acad Sci USA* 2, 85-97.

Creamer, T. P., & Rose, G. D. (1994). Alpha-helix-forming propensities in peptides and proteins. *Proteins* 19, 85-97.

Creighton, T. E. (1990). Protein folding. *Biochem J* 270, 1-16.

Crick, F. H. C. (1953). The packing of alpha-helices: simple coiled-coils. *Acta Cryst* 6, 689-697.

Cregut, D., Civera, C., Macias, M.J., Wallon, G., and Serrano, L. (1999). A tale of two secondary structure elements: when a β-hairpin becomes an α-helix. *J. Mol. Biol.* 292, 389-401.

Dalal, S., Balasubramanian, S., & Regan, L. (1997). Protein alchemy: changing beta-sheet into alpha-helix. *Nat Struct Biol* 4, 548-552.

Dao-pin, S., Nicholson, H., Baase, W. A., Zhang, X. J., Wozniak, J. A., & Matthews, B. W. (1991). Structural and genetic analysis of electrostatic and other interactions in bacteriophage T4 lysozyme. *Ciba Found Symp* 161, 52-62.

De Crescenzo, G., Litowski, J. R., Hodges, R. S. & O'Connor-McCourt, M. D. (2003). Real-time monitoring of the interactions of two-stranded de novo designed coiled-coils: effect of chain length on the kinetic and thermodynamic constants of binding. *Biochemistry* 42, 1754-1763.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry* 29, 7133-7155.

Dill, K. A. (1999). Polymer principles and protein folding. *Protein Sci* 8, 1166-1180.

Dill, K. A., Phillips, A. T., & Rosen, J. B. (1997). Protein structure and energy landscape dependence on sequence using a continuous energy function. *J Comput Biol* 4, 227-239.

Dobson, C. M. (2003). Protein folding and misfolding. *Nature* 426, 884-890.

Doig, A. J., & Baldwin, R. L. (1995). N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci* 4, 1325-1336.

Doig, A. J. (2002). Recent advances in helix-coil theory. *Biophys Chem* 101-102, 281-293.

Dragan, A. I. & Privalov, P. L. (2002). Unfolding of a leucine zipper is not a simple two-state transition. *J Mol Biol* 321, 891-908.

Durr, E., & Jelesarov, I. (2000). Thermodynamic analysis of cavity creating mutations in an engineered leucine zipper and energetics of glycerol-induced coiled coil stabilization. *Biochemistry* 39, 4472-4482.

Eidsness, M. K., Richie, K. A., Burden, A. E., Kurtz, D. M. Jr., Scott, R. A. (1997). Dissecting contributions to the thermostability of Pyrococcus furiosus rubredoxin: beta-sheet chimeras. *Biochemistry* 36, 10406-10413.

Eisenberg, D., & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature* 319, 199-203.

Eisenhaber, F., Imperiale, F., Argos, P., & Frommel, C. (1996). Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods. *Proteins* 25, 157-168.

Engel, J. (1991). Domains in proteins and proteoglycans of the extracellular matrix with functions in assembly and cellular activities. *Int J Biol Macromol* 13, 147-151.

Fairman, R., Shoemaker, K. R., York, E. J., Stewart, J. M., & Baldwin, R. L. (1989). Further studies of the helix dipole model: effects of a free alpha-NH3+ or alpha-COO- group on helix stability. *Proteins* 5, 1-7.

Fairman, R., Chao, H.G., Mueller, L., Lavoie, T.B., Shen, L., Novotny, J., and Matsueda, G.R. (1995). Characterization of a new four-chain coiled-coil: influence of chain length on stability. *Protein Sci.* 4, 1457-1469.

Fauchere, J., & Pliska, V. (1983). Hydrophobic parameters of amino acid side chains from the portioning of N-acetyl-amino acid amides. *Eur J Med Chem* 18, 369-375.

Ferguson, N., & Fersht, A. R. (2003). Early events in protein folding. *Curr Opin Struct Biol* 13, 75-81.

Fields, G. B., Ed. (1997). *Solid-Phase Peptide Synthesis: Methods Enzymol.* New York, Academic Press.

Flory, P. J. (1969). *Statistical Mechanics of Chain Molecules*, Interscience, New York.

Fontenot, J.D., Ball, J.M., Miller, M.A., David, C.M., and Montelaro, R.C. (1991). A survey of potential problems and quality control in peptide synthesis by the fluorenylmethoxycarbonyl procedure. *J. Pep. Res.* 1, 19-25.

Gallo, S. A., Finnegan, C. M., Viard, M., Raviv, Y., Dimitrov, A., Rawat, S. S., Puri, A., Durell, S. & Blumenthal, R. (2003). The HIV Env-mediated fusion reaction. *Biochim Biophys Acta* 1614, 36-50.

Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W., & Kallenbach, N. R. (1991). The helix-coil transition in heterogeneous peptides with specific side-chain interactions: theory and comparison with CD spectral data. *Biopolymers* 31, 1605-1614.

Gasset, M., Baldwin, M. A., Lloyd, D. H., Gabriel, J. M., Holtzman, D. M., Cohen, F., Fletterick, R., & Prusiner, S. B. (1992). Predicted alpha-helical regions of the prion protein when synthesized as peptides form amyloid. *Proc Natl Acad Sci USA* 89, 10940-10944.

Gee, M. A., Heuser, J. E., & Vallee, R. B. (1997). An extended microtubule-binding structure within the dynein motor domain. *Nature* 390, 636-639.

Gomez, J., Hilser, V. J., Xie, D. & Freire, E. (1995). The heat capacity of proteins. *Proteins* 22, 404-412.

Gonzalez, L. Jr., Plecs, J. J., Alber, T. (1996). An engineered allosteric switch in leucine-zipper oligomerization. *Nat Struct Biol* 3, 510-515.Grant, G. A., Ed. (1992). *Synthetic Peptides: A User's Guide.* New York, W. H. Freeman and Company.

Goodman, E. M., & Kim, P. S. (1991). Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide. *Biochemistry* 30, 11615-11620.

Hamada, D., Kuroda, Y., Tanaka, T., & Goto, Y. (1995). High helical propensity of the peptide fragments derived from beta-lactoglobulin, a predominantly beta-sheet protein. *J Mol Biol* 254, 737-746.

Hanson, M., Unger, K. K., Mant, C. T., Hodges, R. S. (1992). Polymer-coated reversed-phase packings with controlled hydrophobic properties. II. Effect on selectivity of peptide separations. *J Chromatogra* 599, 77-85.

Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262, 1401-1407.

Harbury, P. B., Kim, P. S. & Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature* 371, 80-83.

Hermans, J., Anderson, A. G., & Yun, R. H. (1992). Differential helix propensity of small apolar side chains studied by molecular dynamics simulations. *Biochemistry* 31, 5646-5653.

Hill, R. B., Raleigh, D. P., Lombardi, A. & DeGrado, W. F. (2000). De novo design of helical bundles as models for understanding protein folding and function. *Acc Chem Res* 33, 745-754.

Hitchcock-DeGregori, S.E., Song, Y., & Greenfield, N.J. (2002). Functions of tropomyosin's periodic repeats. *Biochemistry* 41, 15036-15044.

Hodges, R. S., Sodek, J., Smillie, L. B., & Jurasek, L. (1972). Tropomyosin: amino acid sequence and coiled-coil structure. *Cold Spring Harbor Symp. Quant. Biol.* 37, 299-310.

Hodges, R.S., Saund, A.K., Chong, P.C.S., St-Pierre, S.A., & Reid, R.E. (1981). Synthetic model for two-stranded α-helical coiled-coils. *J. Biol. Chem.* 256, 1214-1224.

Hodges, R. S., Semchuk, P. D., Taneja, A. K., Kay, C. M., Parker, J. M. & Mant, C. T. (1988). Protein design using model synthetic peptides. *Pept Res* 1, 19-30.

Hodges, R. S., Zhou, N. E., Kay, C. M. & Semchuk, P. D. (1990). Synthetic model proteins: contribution of hydrophobic residues and disulfide bonds to protein stability. *Pept Res* 3, 123-137.

Hodges, R. S. (1992). Unzipping the secret of coiled-coils. *Curr Biol* 2, 122-124.

Hodges, R. S., Zhu, B.-Y., Zhou, N. E., & Mant, C. T. (1994). Reversed-phase chromatography as auseful probe of hydrophobic interactions involved in protein folding and protein stability. *J Chromatogr A* 676, 3-15.

Hodges, R. S. (1996). Boehringer Mannheim award lecture 1995. De novo design of alpha-helical proteins: basic research to medical applications. *Biochem Cell Biol* 74, 133-154.

Holtzer, A. & Skolnick, J. (1988). Application of the augmented theory of alpha-helix-to-random-coil transitions of two-chain, coiled coils to extant data on synthetic, tropomyosin-analog peptides. *Biopolymers* 27, 87-96.

Holtzer, A. & Holtzer, M. E. (1990). Alpha-helix to random-coil transitions of two-chain coiled coils: the use of physical models in treating thermal denaturation equilibria of isolated subsequences of alpha alpha-tropomyosin. *Biopolymers* 30, 1231-1241.

Holtzer, M. E., Lovett, E. G., d'Avignon, D. A. & Holtzer, A. (1997). Thermal unfolding in a GCN4-like leucine zipper: 13C alpha NMR chemical shifts and local unfolding curves. *Biophys J* 73, 1031-1041.

Honig, B., Ray, A., & Levinthal, C. (1976). Conformational flexibility and protein folding: rigid structural fragments connected by flexible joints in subtilisin. *Proc Natl Acad Sci USA* 73, 1974-1978.

Honig, B., & Yang, A.-S. (1995). Free energy balance in protein folding. *Adv Prot Chem* 46, 27-58.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M., & Fersht, A. R. (1990). Strength and co-operativity of contributions of surface salt bridges to protein stability. *J Mol Biol* 216, 1031-1044.

Horovitz, A., Matthews, J. M., Fersht, A. R. (1992). Alpha-helix stability in proteins. II. Factors that influence stability at an internal position. *J Mol Biol* 227, 560-568.

Houston, M. E., Jr., Campbell, A.P., Lix, B., Kay, C.M., Sykes, B.D., and Hodges, R.S. (1996a). Lactam bridges stabilization of alpha-helices: the role of hydrophobicity in controlling dimeric versus monomeric alpha-helices. *Biochemistry* 35, 10041-10050.

Houston, M. E. Jr, Wallace, A., Bianchi, E., Pessi, A., & Hodges, R. S. (1996b). Use of a conformationally restricted secondary structural element to display peptide libraries: a two-stranded alpha-helical coiled-coil stabilized by lactam bridges. *J Mol Biol* 262, 270-282.

Jelesarov, I., & Bosshard, H. R. (1996). Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). *J Mol Biol* 263, 344-358.

Jelesarov, I., Durr, E., Thomas, R. M. & Bosshard, H. R. (1998). Salt effects on hydrophobic interaction and charge screening in the folding of a negatively charged peptide to a coiled coil (leucine zipper). *Biochemistry* 37, 7539-7550.

Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys J* 36, 575-588.

Kammerer, R.A., Schulthess, T., Landwehr, R., Engel, J., Aebi, U., Steinmetetz, M.O. (1998). An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc. Natl. Acad. Sci, USA* 95, 13419-13424.

Kamtekar, S., & Hecht, M. H. (1995). Protein Motifs. 7. The four-helix bundle: what determines a fold? *FASEB J* 9, 1013-1022.

Karplus, M., & Weaver, D. L. (1976). Protein-folding dynamics. *Nature* 260, 404-406.

Karplus, M. (1997). Hydrophobicity regained. *Prot Sci* 6, 1302-1307.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14, 1-63.

Kellis, J.T., Jr., Nyberg, K., and Fersht, A.R. (1989). Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28, 4914-4922.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181, 662-666.

Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science* 295, 1719-1722.

Kohn, W. D., Kay, C. M. & Hodges, R. S. (1995a). Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper. *Prot Sci* 4, 237-250.

Kohn, W. D., Monera, O. D., Kay, C. M. & Hodges, R. S. (1995b). The effects of interhelical electrostatic repulstions between glutamic acid residues in controlling the dimerization and stability of two-stranded alpha-helicals. *J Biol Chem* 270, 25495-25506.

Kohn, W. D., Kay, C. M. & Hodges, R. S. (1997a). Salt effects on protein stability: two-stranded alpha-helical coiled-coils containing inter- or intrahelical ion pairs. *J Mol Biol* 267, 1039-1052.

Kohn, W. D., Mant, C. T., Hodges, R, S. (1997b). Alpha-helical protein assembly motifs. *J Biol Chem* 272, 2583-2586.

Kohn, W. D., Mant, C. T., Hodges, R, S. (1997c). Positional dependence of the effects of negatively charged Glu side-chains on the stability of two-stranded α-helical coiled-coils. *J Pept Sci* 3, 209-223.

Kohn, W. D., Kay, C. M. & Hodges, R. S. (1998d). Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils. *J Mol Biol* 283, 993-1012.

Kohn, W. D. & Hodges, R. S. (1998). De novo design of alpha-helical coiled-coils and bundles: protein models for development of design principles. *Trends Biotech* 16, 379-389.

Kohn, W. D., Kay, C. M. & Hodges, R. S. (1998a). Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils. *J Mol Biol* 283, 993-1012.

Kohn, W. D., Kay, C. M., & Hodges, R. S. (1998b). Effects of lanthanide binding on the stability of *de novo* designed α-helical coiled-coil. *J Pept Res* 51, 9-18.

Kohn, W. D., Kay, C. M., Sykes, B. D., & Hodges, R. S. (1998). Metal ion induced folding of a *de novo* designed coiled-coil peptide. *J Am Chem Soc* 120, 1124-1132.

Kuroda, Y., Hamada, D., Tanaka, T., & Goto, Y. (1996). High helicity of peptide fragments corresponding to beta-strand regions of beta-lactoglobulin observed by 2D-NMR spectroscopy. *Fold Des* 4, 255-263.

Kwok, S.C., Tripet, B., Man, J.H., Chana, M.S., Lavigne, P., Mant, and C.T., Hodges, R.S. (1998a). Structural cassette mutagenesis in a *de novo* designed protein: proof of a novel concept for examining protein folding and stability. *Biopolymers (Pept. Sci.)* 47, 101-123.

Kwok, S.C., Mant, C.T., and Hodges, R.S. (1998b). Effect of α-helical and β-sheet propensities of amino acids on protein stability. *Peptides 1998: Proc. 25th European Peptide Symposium,* p.34-35.

Kwok, S.C., Mant, C.T., and Hodges, R.S. (2001). Using conformationally-restricted lactam-bridged peptides to examine the effects of α-helical propensity of solvent-exposed residues. *Peptides: the wave of the future: Proc. 2nd International and 17th American Peptide Symposium.*

Kwok, S. C., Mant, C. T. & Hodges, R. S. (2002). Importance of secondary structural specificity determinants in protein folding: insertion of a native beta-sheet sequence into an alpha-helical coiled-coil. *Protein Sci* 11, 1519-1531.

Kwok, S. C. & Hodges, R. S. (2003). Clustering of large hydrophobes in the hydrophobic core of two-stranded alpha -helical coiled-coils control protein folding and stability. *J Biol Chem* 278, 35248-35254.

Kwok, S. C., & Hodges, R. S. (2004a). Stabilizing and destabilizing clusters in the hydrophobic core of long two-stranded α-helical coiled-coils. *J Biol Chem*, submitted.

Kwok, S. C., & Hodges, R. S. (2004b). Effect of chain length on coiled-coil stability: decreasing stability with increasing chain length. *Biopolymers (Pept Sci)*, submitted.

Landis, C., Back, N., Homsher, E., & Tobacman, L.S. (1999). Effects of tropomyosin internal deletions on thin filament function. *J Biol Chem* 274, 31279-31285.

Lattman, E.E. and Rose, G.D. (1993). Protein folding-what's the question? *Proc. Natl. Acad. Sci. USA* 90, 439-441.

Lau, S.Y.M., Taneja, A.K., & Hodges, R.S. (1984a). Synthesis of a model protein of defined secondary and quaternary structure; effect of chain length on the stabilization and formation of two-stranded α-helical coiled-coils. *J Biol Chem* 259, 13253-13261.

Lau, S.Y.M., Taneja, A.K., & Hodges, R.S. (1984b). Effects of HPLC solvents and hydrophobic supports on the secondary and quaternary structure of a model protein: Reversed-phase and size-exclusion high performance liquid chromatography. *J Chromatogr* 317, 129-140.

Lavigne, P., Kondejewski, L. H., Houston, M. E., Jr., Sonnichsen, F. D., Lix, B., Skyes, B. D., Hodges, R. S. & Kay, C. M. (1995). Preferential heterodimeric parallel coiled-coil formation by synthetic Max and c-Myc leucine zippers: a description of putative electrostatic interactions responsible for the specificity of heterodimerization. *J Mol Biol* 254, 505-520.

Lazaridis, T., Archontis, G., & Karplus, M. (1995). Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Adv Prot Chem* 47, 231-306.

Lee, D.L., Lavigne, P., and Hodges, R.S. (2001). Are trigger sequences essential in the folding of two-stranded α-helical coiled-coils? *J. Mol. Biol.* 306, 539-553.

Lehmann, M., Pasamontes, L., Lassen, S.F., Wyss, M. (2000). The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta.* 1543, 408-415.

Levinthal, C. (1969). Are there pathways for protein folding? *J. Chim. Phys.* 65, 44-45.

Li, S. C. & Deber, C. M. (1994). A measure of helical propensity for amino acids in membrane environments. *Nat Struct Biol* 1, 368-373.

Li, S. C., Goto, N. K., Williams, K. A., Deber, C. M. (1996). Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc Natl Acad Sci USA* 93, 6676-6681.

Litowski, J. R. & Hodges, R. S. (2001). Designing heterodimeric two-stranded alpha-helical coiled-coils: the effect of chain length on protein folding, stability and specificity. *J Pept Res* 58, 477-492.

Litowski, J. R., & Hodges, R. S. (2002). Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. *J Biol Chem* 277, 37272-37279.

Liu, J., Cao, W., & Lu, M. (2002). Core side-chain packing and backbone conformation in Lpp-56 coiled-coil mutants. *J Mol Biol* 318, 877-888.

Lockhart, D. J., & Kim, P. S. (1993). Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* 260, 198-202.

Lu, M., Shu, W., Ji, H., Spek, E., Wang, L., & Kallenbach, N. R. (1999). Helix capping in the GCN4 leucine zipper. *J Mol Biol* 288, 743-752.

Lu, S., & Hodges, R. S. (2004). Defining the minimum size of a hydrophobic cluster in two-stranded α-helical coiled-coils: effects on protein stability. *Prot Sci* in press.

Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162-1164.

Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem Sci.* 21, 375-382.

Lyu, P.C., Liff, M.I., Marky, L.A., and Kallenbach, N.R. (1990) Side chain contributions to the stability of alpha-helical structure in peptides. *Science* 250, 669-673.

Lyu, P. C., Gans, P. J., & Kallenbach, N. R. (1992). Energetic contribution of solvent-exposed ion pairs to alpha-helix structure. *J Mol Biol* 223, 343-350.

Makhatadze, G. I., & Privalov, P. L. (1990). Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. *J Mol Biol* 213, 375-384.

Makhatadze, G. I. & Privalov, P. L. (1995). Energetics of protein structure. *Adv Protein Chem* 47, 307-425.

Mant, C. T., Chao, H., & Hodges, R. S. (1997a). Effect of mobile phase on the oligomerization state of alpha-helical coiled-coil peptides during high performance size-exclusion chromatography. *J Chromatogr A* 791, 85-98.

Mant, C. T., Kondejewski, L. H., Cachia, P. J., Monera, O. D. & Hodges, R. S. (1997b). Analysis of synthetic peptides by high-performance liquid chromatography. *Methods Enzymol* 289, 426-469.

Mant, C.T, Kondejewski, L.H., Cachia, P.J., Monera, O.D., and Hodges, R.S. (1997c). Practical aspects of analysis of synthetic peptides by high-performance liquid chromatography. *Meth. Enzymol.* 289, 426-469.

Mant, C. T., Litowski, J. R., & Hodges, R. S. (1998). Hydrophilic interaction/cation exchange chromatography for separation of amphipathic alpha-helical peptides. *J Chromatogr A* 816, 65-78.

Mant, C. T., & Hodges, R. S. (2002). Analytical HPLC of Peptides. *HPLC of Biological Macromolecules.* (Eds. Gooding, K. M., & Regnier, F. E.). New York, Marcel Dekker, Inc., pp.. 433-511.

Makhatadze, G. I., & Privalov, P. L. (1995). Energetics of protein structure *Adv Protein Chem* 47, 307-425.

Matthews, B. W. (1995). Studies on protein stability with T4 lysozyme. *Adv Protein Chem* 46, 249-278.

McClain, D. L., Binfet, J. P. & Oakley, M. G. (2001). Evaluation of the energetic contribution of interhelical Coulombic interactions for coiled coil helix orientation specificity. *J Mol Biol* 313, 371-383.

Meizi, M. (1998). Chameleon sequences in the PDB. Protein *Eng.* 11, 411-414.

Merrifield, B. (1997). Concept and early development of solid-phase peptide synthesis. *Methods Enzymol* 289, 3-13.

Merutka, G., & Stellwagen, E. (1990a). Positional independence and additivity of amino acid replacements on helix stability in monomeric peptides. *Biochemistry* 29, 894-898.

Merutka, G., Lipton, W., Shalongo, W., Park, S. H., & Stellwagen, E. (1990b). Effect of central-residue replacements on the helical stability of a monomeric peptide. *Biochemistry* 29, 7511-7515.

Mezei, M. (1998). Chameleon sequences in the PDB. *Protein Eng* 11, 411-414.

Micklatcher, C. & Chmielewski, J. (1999). Helical peptide and protein design. *Curr Opin Chem Biol* 3, 724-729.

Minor, D.L., Jr., and Kim, P.S. (1994). Measurement of the beta-sheet-forming propensities of amino acids. *Nature* 367, 660-663.

Minor, D. L., Jr. & Kim, P. S. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature* 380, 730-734.

Misra, G. P., & Wong, C. F. (1997). Predicting helical segments in proteins by a helix-coil transition theory with parameters derived from a structural database of proteins. *Proteins* 28, 344-359.

Monera, O. D., Zhou, N. E., Kay, C. M., & Hodges, R. S. (1993). Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *J Biol Chem* 268, 19218-19227.

Monera, O. D., Kay, C. M., Hodges, R. S. (1994a). Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci* 3, 1984-1991.

Monera, O. D., Kay, C. M., Hodges, R. S. (1994b). Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry* 33, 3862-3871.

Monera, O.D., Sereda, T.J., Zhou, N.E., Kay, C.M., and Hodges, R.S. (1995). Relationship of side-chain hydrophobicity and α-helical propensity on the stability of single-stranded amphipathic α-helix. *J. Pep. Res.* 1, 312-329.

Monera, O. D., Zhou, N. E., Lavigne, P., Kay, C. M., Hodges, R. S. (1996a). Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J Biol Chem* 271, 3995-4001.

Monera, O. D., Sonnichsen, F. D., Hicks, L., Kay, C. M., & Hodges, R. S. (1996b). The relative positions of alanine residues in the hydrophobic core control the formation of two-stranded or four-stranded alpha-helical coiled-coils. *Protein Eng* 9, 353-363.

Monne, M., Hermansson, M., and von Heijne, G. (1999). A turn propensity scale for transmembrane helices. *J. Mol. Biol.* 288, 141-145.

Munoz, V., Cronet, P., Lopez-Hernandez, E., & Serrano, L. (1996). Analysis of the effect of local interactions on protein stability. *Fold Des* 3, 167-178.

Murphy, K. P., Privalov, P. L., & Gill, S. J. (1990). Common features of protein unfolding and dissolution of hydrophobic compounds. *Science* 247, 559-561.

Myers, J. K., Pace, C. N., & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Prot Sci* 4, 2138-2148.

Nedwidek, M. N., & Hecht, M. H. (1997). Minimized protein structures: a little goes a long way. *Proc Natl Acad Sci USA* 94, 10010-10011.

Novabiochem. (2003). *Novabiochem Catalog and Peptide Synthesis Handbook 2003/2004*, Novabiochem, San Diego, CA.

Nozaki, Y., & Tanford, C. (1971). The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J Biol Chem 246*, 2211-2217.

Ohmura, T., Ueda, T., Hashimoto, and Y., Imoto, T. (2001). Tolerance of point substitution of methionine for isoleucine in hen egg white lysozyme. *Protein Eng.* 14, 421-425.

O'Neil, K. T. & DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250, 646-651.

O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254, 539-544.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131, 266-280.

Pace, C. N., Shirley, B. A., McNutt, M., & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J* 10, 75-83.

Pace, C. N. (2001). Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry* 40, 310-313.

Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., & Baldwin, R. L. (1990). Relative helix-forming tendencies of nonpolar amino acids. *Nature* 344, 268-270.

Padmanabhan, S., & Baldwin, R. L. (1994). Helix-stabilizing interaction between tyrosine and leucine or valine when the spacing is i, i + 4. *J Mol Biol* 241, 706-713.

Pan, K. M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R. J., Cohen, F. E., & Prusiner, S. B. (1993).

Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA* 23, 10962-10966.

Park, S. H., Shalongo, W., Stellwagen, E. (1993). Residue helix parameters obtained from dichroic analysis of peptides of defined sequence. *Biochemistry* 32, 7048-7053.

Paulucci, A. A., Hicks, L., Machado, A., Miranda, M. T., Kay, C. M., & Farah, C. (2002). Specific sequences determine the stability and cooperativity of folding of the C-terminal half of tropomyosin. *J Biol Chem* 277, 39574-39584.

Penel, S., and Doig, A.J. (2001). Rotamer strain energy in protein helices quantification of a major force opposing protein folding. *J. Mol. Biol.* 305, 961-968.

Perry, S.V. (2001). Vertebrate tropomyosin: distribution, properties and function. *J Muscle Res Cell Motil* 22, 5-49.

Pertsemlidis, A., Saxena, A. M., Soper, A. K., Head-Gordon, T., & Glaeser, R. M. (1996). Direct evidence for modified solvent structure within the hydration shell of a hydrophobic amino acid. *Proc Natl Acad Sci USA* 93, 10769-10774.

Ptitsyn, O. B., Denesyuk, A. I., Finkelstein, A. V., Lim, V. I. (1973). *FEBS Lett* 34, 55-37.

Presta, L. G. & Rose, G. D. (1988). Helix signals in proteins. *Science* 240, 1632-1641.

Privalov, P. L., & Potekhin, S. A. (1986). Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol* 131, 4-51.

Prusiner, S. B. (1997). Prion diseases and the BSE crisis. *Science* 278, 245-251.

Prusiner, S. B. (1998). Prions. *Proc Natl Acad Sci USA* 95, 13363-13383.

Qian, H. (1994). A thermodynamic model for the helix-coil transition coupled to dimerization of short coiled-coil peptides. *Biophys J* 67, 349-355.

Ramirez-Alvarado, M., Serrano, L., & Blanco, F. J. (1997). Conformational analysis of peptides corresponding to all the secondary structure elements of protein L B1 domain: secondary structure propensities are not conserved in proteins with the same fold. *Protein Sci* 6, 162-174.

Regan, L. & Jackson, S. E. (2003). Engineering and design. Protein design: theory and practice. *Curr Opin Struct Biol* 13, 479-481.

Reymond, M. T., Huo, S., Duggan, B., Wright, P. E., & Dyson H. J. (1997a). Contribution of increased length and intact capping sequences to the conformational preference for helix in a 31-residue peptide from the C terminus of myohemerythrin. *Biochemistry* 36, 5234-5244.

Reymond, M. T., Merutka, G., Dyson, H. J., Wright, P. E. (1997b). Folding propensities of peptide fragments of myoglobin. *Protein Sci* 3, 706-716.

Rohl, C. A., Chakrabartty, A., & Baldwin, R. L. (1996). Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci* 5, 2623-2637.

Rost, B., & Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 19, 55-72.

Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27, 8063-8068.

Schellman, J. A. (1955). The stability of hydrogen-bonded peptide structures in aqueous solution. *C R Trav Lab Carlsberg Ser Chim* 29, 230-259.

Scheraga, H. A., Vila, J. A., Ripoll, D. R. (2002). Helix-coil transitions re-visited. *Biophys Chem* 101-102, 255-265.

Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M., Bolen, D. W. (1991a). Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. *Proc Natl Acad Sci USA* 88, 2854-2858.

Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M., & Baldwin, R. L. (1991b). Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* 31, 1463-1470.

Scholtz, J. M. & Baldwin, R. L. (1992). The mechanism of alpha-helix formation by peptides. *Annu Rev Biophys Biomol Struct* 21, 95-118.

Scholtz, J. M., Qian, H., Robbins, V. H. & Baldwin, R. L. (1993). The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* 32, 9668-9676.

Schultz, C. N., & Warshel, A. (2001). What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* 44, 400-417.

Sellers, J. R. (2000). Myosins: a diverse superfamily. *Biochim Biophys Acta* 17, 3-22.

Seo J., & Cohen, C. (1993). Pitch diversity in alpha-helical coiled coils. *Proteins* 15, 223-234.

Sereda, T. J., Mant, C. T., Quinn, A. M. & Hodges, R. S. (1993). Effect of the alpha-amino group on peptide retention behaviour in reversed-phase chromatography. Determination of the pK(a) values of the alpha-amino group of 19 different N-terminal amino acid residues. *J Chromatogr* 646, 17-30.

Sereda, T. J., Mant, C. T., Sonnichsen, F. D., & Hodges, R. S. (1994). Reversed-phase chromatography of synthetic amphipathic alpha-helical peptides as a model for ligand/receptor interactions. Effect of changing hydrophobic environment on the relative hydrophilicity/hydrophobicity of amino acid side-chains. *J Chromatogr A* 676, 139-153.

Sereda, T. J., Mant, C. T., & Hodges, R. S. (1995). Selectivity due to conformational differences between helical and non-helical peptides in reversed-phase chromatography. *J Chromatogr A* 695, 205-221.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M., & Fersht, A. R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* 29, 9343-9352.

Serrano, L., Sancho, J., Hirshberg, M., & Fersht, A. R. (1992). Alpha-helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. *J Mol Biol* 227, 544-559.

Serrano, L., Day, A. G., & Fersht, A. R. (1993). Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J Mol Biol* 233, 305-312.

Serrano, L., & Fersht A. R. (1989). Capping and alpha-helix stability. *Nature* 342, 296-299.

Serrano, L. (2000). The relationship between sequence and structure in elementary folding units. *Adv Prot Chem* 53, 49-85.

Shea, J. E., & Brooks, C. L. 3rd. (2001). From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem* 52, 499-535.

Sheetz, M. P. (1999). Motor and cargo interactions. *Eur J Biochem* 262, 19-25.

Shortle, D. (1989). Probing the determinants of protein folding and stability with amino acid substitutions. *J Biol Chem* 264, 5315-5318.

Shortle, D., & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293, 487-490.

Singh, A., & Hitchcock-DeGregori, S. E. (2003). Local destabilization of the tropomyosin coiled-coil gives the molecular flexibility required for actin binding. *Biochemistry* 42, 14114-14121.

Skolnick, J. (1983). Effect of Loop entropy on the helix-coil transition of a α-helical, two-chain, coiled-coils. *Macromolecules* 16, 1069-1083.

Skolnick, J., & Holtzer, A. (1986). Alpha-helix-to-random-coil transitions of two-chain, coiled coils: a theoretical model for the "pretransition" in cysteine-190-cross-linked tropomyosin. *Biochemistry* 25, 6192-6202.

Smith, C. K., Withka, J. M., Regan, L. (1994). A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* 33, 5510-5517.

Sodek, J., Hodges, R. S., Smillie, L. B., & Jurasek, L. (1972). Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure. *Proc Natl Acad Sci USA* 69, 3800-3804.

Sonnichsen, F. D., Van Eyk, J. E., Hodges, R. S. & Sykes, B. D. (1992). Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry* 31, 8790-8798.

Sosnick, T. R., Jackson, S., Wilk, R. R., Englander, S. W., DeGrado, W. F. (1996). The role of helix formation in the folding of a fully alpha-helical coiled coil. *Proteins* 24, 427-432.

Stellwagen, E., Park, S. H., Shalongo, W. & Jain, A. (1992). The contribution of residue ion pairs to the helical stability of a model peptide. *Biopolymers* 32, 1193-1200.

Su, J. Y., Hodges, R. S. & Kay, C. M. (1994). Effect of chain length on the formation and stability of synthetic alpha-helical coiled coils. *Biochemistry* 33, 15501-15510.

Suarez, M.C., Lehrer, S.S., & Silva, J.L. (2001). Local heterogeneity in the pressure denaturation of the coiled-coil tropomyosin because of subdomain folding units. *Biochemistry* 40, 1300-1307.

Tidor, B. (1994). Helix-capping interaction in lambda Cro protein: a free energy simulation analysis. *Proteins* 19, 310-323.

Timpl, R., & Brown, J. C. (1994). The laminins. *Matrix Biol* 14, 275-281.

Tripet, B., Yu, L., Bautista, D. L., Wong, W. Y., Irvin, R. T., & Hodges, R. S. (1996). Engineering a de novo-designed coiled-coil heterodimerization domain off the rapid detection, purification and characterization of recombinantly expressed peptides and proteins. 9, 1029-1042.

Tripet, B., Van Eyk, J. E., & Hodges, R. S. (1997). Mapping of a second actin-tropomyosin and a second troponin C binding site within the C terminus of

troponin I, and their importance in the Ca2+-dependent regulation of muscle contraction. *J Mol Biol* 271, 728-750.

Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T. & Hodges, R. S. (2000). Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". *J Mol Biol* 300, 377-402.

Tripet, B. & Hodges, R. S. (2001). *STABLECOIL: an algorithm designed to predict the location and relative stability of coiled-coils in native protein sequences.* Proceedings of the Seventeenth American Peptide Symposium (Lebh, M. & Houghten, R. A., Eds.), American Peptide Society, San Diego.

Tripet, B., & Hodges, R. S. (2002). Helix capping interactions stabilize the N-terminus of the kinesin neck coiled-coil. *J Struct Biol* 137, 220-235.

Vallee, R. B., & Gee, M. A. (1998). Make room for dynein. *Trends Cell Biol* 8, 490-494.

Vendruscolo, M., Paci, E., Dobson, C. M., & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature* 409, 641-645.

Wagschal, K., Tripet, B. & Hodges, R. S. (1999a). De novo design of a model peptide sequence to examine the effects of single amino acid substitutions in the hydrophobic core on both stability and oligomerization state of coiled-coils. *J Mol Biol* 285, 785-803.

Wagschal, K., Tripet, B., Lavigne, P., Mant, C. & Hodges, R. S. (1999b). The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein Sci* 8, 2312-2329.

Wan, W. Y., & Milner-White, E. J. (1999a). A natural grouping of motifs with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: their occurrence at alpha-helical N termini and in other situations. *J Mol Biol* 286, 1633-1649.

Wan, W. Y., & Milner-White, E. J. (1999b). A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at alpha-helical N termini and in other situations. *J Mol Biol* 286, 1651-1662.

Watanabe, S., Takada, A., Watanabe, T., Ito, H., Kida, H. & Kawaoka, Y. (2000). Functional importance of the coiled-coil of the Ebola virus glycoprotein. *J Virol* 74, 10194-10201.

Weber, P. C., & Salemme, F. R. (2003). Applications of calorimetric methods to drug discovery and the study of protein interactions. *Curr Opin Struct Biol* 13, 115-121.

Weiss, A., Schiaffino, S., & Leinwand, L. A. (1999). Comparative sequence analysis of the complete human sarcomeric myosin heavy chain family: implications for functional diversity. *J Mol Biol* 290, 61-75.

Wishart, D.S., Boyko, R.F., Willard, L., Richards, F.M., and Sykes, B.D. (1994). SEQSEE: a comprehensive program suite for protein sequence analysis. *Comput. Appl. Biosci.* 10, 121-132.

Wishart, D. S., Fortin S., Woloschuk, D.R., Wong, W., Rosborough, T., van Domselaar, G., Schaeffer, J., and Szafron, D. (1997). A platform-independent graphical user interface for SEQSEE and XALIGN. *Comput. Appl. Biosci.* 13, 561-562.

Wojcik, J., Altmann, K. H., Vasquez, M., & Scheraga, H. A. (1990). Helix-coil stability constants for the naturally occurring amino acids in water. XXIII. Proline parameters from random poly (hydroxybutylglutamine-co-L-proline). *Biopolymers* 30, 107-120.

Wolf, E., Kim, P. S. & Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci* 6, 1179-1189.

Xie, D., & Freire, E. (1994). Structure based prediction of protein folding intermediates. *J Mol Biol* 242, 62-80.

Xiong, H., Buckwalter, B. L., Shieh, H-M., Hecht, M. H. (1995). Periodicity of polar and nonpolar amino acids is the major determinant of secondary

structure in self-assembling oligomeric peptides. *Proc Natl Acad USA* 92, 6349-6353.

Xu, J., Baase, W. A., Baldwin, E., & Matthews, B. W. (1998). The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci* 7, 158-177.

Yang, J., Spek, E. J., Gong, Y., Zhou, H. & Kallenbach, N. R. (1997). The role of context on alpha-helix stabilization: host-guest analysis in a mixed background peptide model. *Protein Sci* 6, 1264-1272.

Yao, J., Chung, J., Eliezer, D., Wright, P. E., & Dyson, H. J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* 40, 3561-3571.

Yon, J. M. (2002). Protein folding in the post-genomic era. *J Cell Mol Med* 6, 307-327.

Yu, Y., Monera, O.D., Hodges, R.S., and Privalov, P.L. (1996). Ion-pairs significantly stabilize coiled-coils in the absence of electrolyte. *J Mol Biol* 255, 367-372.

Yu, B.Y., Lavigne, P., Kay, C.M., Hodges, R.S., and Privalov, P.L. (1999). Contribution of translational and rotational entropy to the unfolding of a dimeric coiled-coil. *J Phys Chem* 103, 2270-2278.

Zaman, M., Berry, R. S. & Sosnick, T. R. (2002). Entropic Benefit of a cross-link in protein association. *Proteins* 48, 341-351.

Zhang, X. J., Baase, W. A., Matthews, B. W. (1992). Multiple alanine replacements within alpha-helix 126-134 of T4 lysozyme have independent, additive effects on both structure and stability. *Protein Sci* 1, 761-776.

Zhou, H. X., Lyu, P., Wemmer, D. E., & Kallenbach, N. R. (1994d). Alpha helix capping in synthetic model peptides by reciprocal side chain-main chain interactions: evidence for an N terminal "capping box". *Proteins* 18, 1-7.

Zhou, H. X., & Dong, F. (2003). Electrostatic contributions to the stability of a thermophilic cold shock protein. *Biophys J* 84, 2216-2222.

Zhou, N. E., Mant, C. T., & Hodges, R. S. (1990). Effect of preferred binding domains on peptide retention behavior in reversed-phase chromatography: amphipathic alpha-helices. *Pept Res* 3, 8-20.

Zhou, N. E., Zhu, B. Y., Kay, C. M. & Hodges, R. S. (1992a). The two-stranded α-helical coiled-coil is an ideal model for studying protein stability and subunit interactions. *Biopolymers* 32, 419-426.

Zhou, N. E., Kay, C. M. & Hodges, R.S. (1992b). Synthetic model proteins: positional effects of interchain hydrophobic interactions on stability of two-stranded alpha-helical coiled-coils. *J Biol Chem* 267, 2664-2670.

Zhou, N. E., Kay, C. M. & Hodges, R. S. (1992c). Synthetic model proteins: the relative contribution of leucine residues at the nonequivalent positions of the 3-4 hydrophobic repeat to the stability of the two-stranded alpha-helical coiled-coil. *Biochemistry* 31, 5739-5746.

Zhou, N. E., Kay, C. M., & Hodges, R. S. (1993a). Disulfide bond contribution to protein stability: positional effects of substitution in the hydrophobic core of the two-A -stranded alpha-helical coiled-coil. *Biochemistry* 32, 3178-3187.

Zhou, N. E., Kay, C. M., Sykes, B. D. & Hodges, R. S. (1993b). A single-stranded amphipathic alpha-helix in aqueous solution: design, structural characterization, and its application for determining alpha-helical propensities of amino acids. *Biochemistry* 32, 6190-6197.

Zhou, N. E., Zhu, B.-Y., Kay, C. M. & Hodges, R. S. (1993c). Importance of intra-chain ionic interactions in stabilizing alpha-helices in proteins. In *Peptides: biology and chemistry (Proceedings of the 1992 Chinese Peptide Symposium)* (Zhang, Y.-S., ed.), pp. 217-220. Escom Science Publishers, Leiden.

Zhou, N. E., Monera, O. D., Kay, C. M., & Hodges, R. S. (1994a). Alpha-helical propensities of amino acids in the hydrophobic face of an amphipathic alpha-helix. *Prot & Pept Lett* 1, 114-119.

Zhou, N. E., Kay, C. M., & Hodges, R. S. (1994b). The role of interhelical ionic interactions in controlling protein folding and stability. De novo designed synthetic two-stranded alpha-helical coiled-coils. *J Mol Biol* 237, 500-512.

Zhou, N. E., Kay, C. M., & Hodges, R. S. (1994c). The net energetic contribution of interhelical electrostatic attractions to coiled-coil stability. *Prot Engr* 7, 1365-1372.

Zhu, B. Y., Zhou, N. E., Semchuk, P. D., Kay, C. M., Hodges, R. S. (1992). Design synthesis and structural characterization of model heterodimeric coiled-coil proteins. *Int J Pept Prot Res* 40, 171-179.

Zhu, B. Y., Zhou, N. E., Kay, C. M., Hodges, R. S. (1993). Packing and hydrophobicity effects on protein folding and stability: effects of beta-branched amino acids, valine and isoleucine, on the formation and stability of two-stranded alpha-helical coiled coils/leucine zippers. *Protein Sci* 2, 383-394.

CURRICULUM VITAE

Name: Stanley C. Kwok

Work Address: Department of Biochemistry and Molecular Genetics
Biomedical Research Building Room 451
The University of Colorado Health Sciences Center
Denver, Coloardo, USA. 80262.

Ph. (303)-315-7344. Fax. (303)-315-1153

Home Address: 1855 San Miguel Road
Victoria, B.C., Canada, V8N 2G9

Ph. 250-721-2350

E-mail: Stan.kwok@uchsc.edu

RESEARCH PROJECTS:

I. *De novo* design, liquid- and solid-phase synthesis, purification and biophysical characterization of peptide and derivatives
II. Understanding the folding and stability of α-helical coiled-coils
III. Investigation of the folding of determinants of short peptide sequences, hydrophobic clustering, chain length, α-helical and β-sheet propensities.

ACADEMIC HISTORY:

2001-present	Transferred to Robert S. Hodges laboratory in University of Colorado Health Sciences Center Denver, Colorado, for completion in doctorate thesis study. (Expected date of completion, Apr '04)
1999	Completion of Doctorate Candidacy Examination
1997-2001	Joined Robert S. Hodges laboratory for graduate studies in the department of biochemistry in University of Alberta, Canada towards a Ph.D. degree
1992-1996	Completion of B. Sc. in Biochemistry in University of Victoria, British Columbia, Canada

Graduate Courses Grade Point Average: 8.3/9.0

HONORS AND AWARDS:

2003	18th American Peptide Symposium Travel Award
1998-2002	Alberta Heritage Foundation for Medical Research Scholarship

1998-2000	National Sciences and Engineering Research Council of Canada Post Graduate Scholarship
1999	16th American Peptide Symposium Travel Award
1998-2000	Walters John Tuition Scholarship
1997	University of Alberta Larry B. Smillie Protein Chemistry Award
1992-1995	Canada Scholarship (Undergraduate, Government of Canada)

PUBLICATIONS

Kwok, S.C., Tripet, B., Man, J.H., Chana, M.S., Lavigne, P., Mant, C.T., and Hodges. R.S. Structural Cassette Mutagenesis in a *de novo* designed protein: Proof of a Novel Concept for Examining Protein Folding and Stability. (1998). *Biopolymers (Pept. Sci.)* 47, 101-119.

Kwok, S.C., Mant, C.T. and Hodges, R.S. Importance of secondary structural specificitifiy determinants in proteins folding. (2002) insertion of a native beta-sheet sequence into an alpha-helical coiled-coil. *Protein Sci.* 11, 1519-1531.

Kwok, S.C., and Hodges, R.S. Clustering of large hydrophobes in the hydrophobic core of two-stranded alpha -helical coiled-coils control protein folding and stability. (2003). *J Biol. Chem.* 278, 35248-35354.

Kwok, S.C., and Hodges, R.S. Stability of two-stranded α-helical coiled coils: context dependent effects of leucine to alanine substitutions adjacent to a destabilzing cluster. *J Biol Chem.* Accepted.

Kwok, S.C., and Hodges, R.S. Effect of chain length on coiled-coil stability: decreasing stability with increasing chain length. *Biopolymers.* Waiting for submission.

POSTER PRESENTATIONS

Kwok, S.C., Mant, C.T., and Hodges, R.S. Effects of α-helical and β-sheet propensities of amino acids on protein stability. Proceedings of the 25th European Peptide Symposium, Budapest, Hungary. 1998.

Kwok, S.C., Mant, C.T., and Hodges, R.S. Using Conformationally-restricted lactam-bridged peptides to examine the effects of α-helical propensity of solvent-exposed residues., San Diego, U.S.A., 2001.

Lu, S.M., Kwok, S.C., and Hodges, R. S. Importance of hydrophobic clusters on protein stability and folding. 18th American Peptide Symposium, Boston, U.S.A., 2003.

INVITED LECTURE

Kwok, S.C., and Hodges, R.S. Balancing enthalpic stabilization and entropy: chain length effects in coiled-coil formation. 18th American Peptide Symposium, Boston, U.S.A., 2003 (Invited speaker).

PROFESSIONAL/WORK EXPERIENCE:

1. University of Alberta/University of Colorado

Title. Ph. D. Graduate Student (1997-Present)

My doctorate thesis focused primarily on the studying the relationship between amino acid sequence and protein folding. Specifically, this involved carrying out knowledge-based *de novo* peptide design, systematic purification and characterization of peptides to study protein interactions that affect folding and stability. Currently, I am completing my thesis writing and involved in a collaborative research project to use peptide approach examine interactions between progesterone receptor and its ligands.

2. SynPhar Laboratories
Edmonton, Alberta, Canada
Title. Student Researcher (1994)

In a four-month summer work term, I developed a mico titer plate minimal inhibitory concentration (MIC) assay to screen the efficacies of potential anti-microbial and anti-fungal compound. This new protocol replaced the previous assay which is agar-based.

TECHNICAL EXPERIENCE

Peptide synthesis: solid- and liquid-phase technologies, fragment condensation, Boc-, Fmoc, and DIC chemistry

Chromatography: Reversed-phase HPLC (Agilent, Beckman and Microtek), size-exclusion, affinity, and ion-exchange.

Molecular Biology: Native and SDS-PAGE, western blotting, capillary electrophoresis.

Biophysical Techniques: Amino acid analysis, Electrospray/MALDI mass spectroscopy (FISON, Mariner,) circular dichroism spectroscopy, sedimentation velocity and sedimentation equilibrium, differential scanning and isothermal titration calorimetry.